The Enzymology of Sugar Metabolism in Sulfolobus solfataricus

Nunn, Charlotte

Award date:
2010

Awarding institution:
University of Bath

The enzymology of sugar metabolism in *Sulfolobus solfataricus*

Charlotte Elizabeth Mary Nunn

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

July 2010

Acknowledgements

I thank my supervisors Dr. David Hough and Prof. Michael Danson for their continued support throughout my PhD. They have at all times, encouraged me to pursue my ideas and take an independent approach to my research while always being there to call on for advice.

I thank Prof. Roy Daniel for giving me the opportunity to undertake a research project in his laboratory at the University of Waikato, and to both he and Prof. Roberta Farrell for their continued help and encouragement.

I would also like to acknowledge all the members of Lab 1.33 who have provided help, support and friendship over the years, but especially Dr Karl Payne and Dr Winnie Wu.

To my husband Collin who has put up with my early mornings and late nights in the lab and my general work obsession for four years, I can only say that I could not have done it without you. I am also very grateful to my parents and family for being so understanding and supportive.

I would like to dedicate this thesis to my Dad who passed away on the 28th June 2009, who I miss.

Contents

Abstract

Metabolism encompasses the biochemical basis of life, but even in well studied model organisms our knowledge of metabolism is incomplete (Downs, 2006). Sugars represent a major source of carbon supporting heterotrophic growth in Archaea and are oxidized via a conserved set of central metabolic pathways, including modified versions of the Entner-Doudoroff (ED) pathway.

Sulfolobus solfataricus is a hyperthermophilic archaeon from the kingdom crenarchaeota. It is a strict aerobe that grows optimally at 80-85°C, pH 2-4, and utilises numerous carbon sources including the four most-commonly occurring sugars in nature, D-glucose, D-galactose, D-xylose and L-arabinose. In *S. solfataricus*, glucose and galactose are metabolised through the non-phosphorylative ED pathway in which a single set of enzymes are utilised for the catabolism of both sugars, leading to the proposal of a metabolically promiscuous pathway. The first section of this pathway includes three enzymes, glucose dehydrogenase (GDH), gluconate dehydratase (GD) and 2-keto-3-deoxy-gluconate (KDG) aldolase, two of which, GDH and KDG aldolase, are found to have activity with pentose as well as hexose sugars, although the third enzyme, GD, is hexose specific. KDG-aldolase cleaves hexose substrates into pyruvate and glyceraldehyde, and pentoses into pyruvate and the C-2 compound glycoaldehyde. Glyceraldehyde is converted into a second molecule of pyruvate, but the fate of glycoaldehyde is currently unknown.

In this thesis, the complete pathway for D-xylose and L-arabinose catabolism is elucidated. Through the detection of enzyme activities in the native organism and subsequent characterisation of recombinantly produced enzymes, the results presented show that a separate C5 dehydratase exists, completing the first part of a non-phosphorylative C5 pathway, and that the subsequent enzymes required for glycolaldehyde metabolism (glycolaldehyde oxidoreductase, glycolate oxidase and malate synthase) are present and active during growth on C5 sugars.

In addition to the metabolic investigations, rational design and site-directed mutagenesis were used to explore the structural basis of substrate promiscuity of glucose dehydrogenase; in doing so, the enzyme was successfully engineered to oxidise the sugar D-ribose while retaining its activity with its natural substrates. These same techniques have

been used to explore the relationship between structure, stability and catalytic activity of the *S. solfataricus* malate synthase, and the data have been fitted to a quantitative kinetic model that explores how temperature influences the activity of an enzyme.

List of abbreviations

DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DCPIP	2,6-dichlorophenolindophenol
ED	Entner-Doudoroff
EM	Embden-Meyerhof
KDGlu	2-keto-3-deoxy-gluconate
KDGal	2-keto-3-deoxy-galactonate
DOP	2,5 dioxopentanoate
GDH	Glucose dehydrogenase
GD	Gluconate dehydratase
XD	Xylonate dehydratase
MS	Malate synthase
GR	Glyoxylate reductase
MDR	Medium chain dehydrogenase / reductase

Chapter 1: General Introduction

1.1. Archaea and extremophiles

The proposal for a third domain of life, distinct from that of bacteria and eukarya, was made in 1990 by Carl Woese and colleagues (Woese *et al*, 1990) based on 16S rRNA sequences. The wealth of sequencing information now available has made this a universally accepted representation of life, despite some earlier contention over its phylogenetic basis (Mayr, 1990). Archaea represent the majority of organisms found in extreme environments, often thriving in conditions that are unfeasible for other organisms even to survive. These extremophilic organisms are generally categorised by their ability to grow in environmental extremes of temperature (-2 – 15°C or 60 – 100°C), salinity (2-5M NaCl) or pH (<4 or >9) (Hough & Danson, 1999), with the molecular basis of this adaptation still a major focus of research.

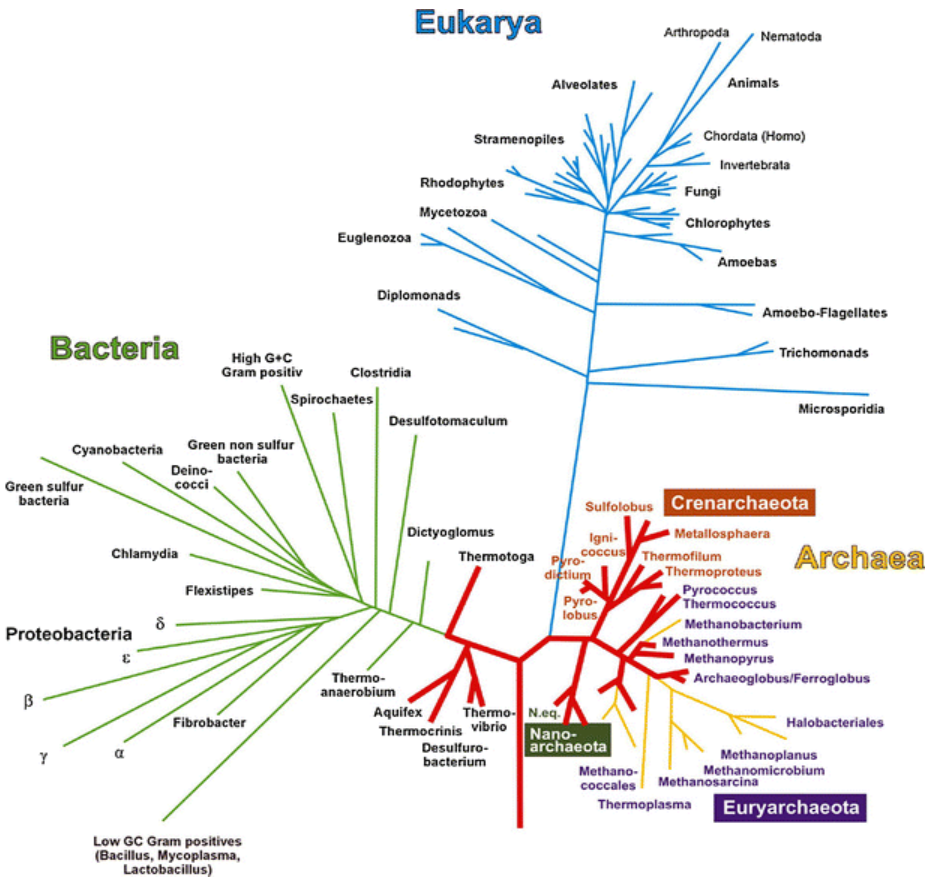


Fig. 1.01. The position of the Archaea and *Sulfolobus* in the phylogenetic tree of life. Taken from Stetter (2006).

1.2. Carbon metabolism

Metabolism encompasses the biochemical basis of life. It is, however, accepted that even in well studied model organisms our knowledge of metabolism is incomplete (Downs, 2006). Sugars represent a major source of carbon supporting heterotrophic growth in Bacteria, Archaea and Eukaryota and are oxidized via a conserved set of central metabolic pathways (Verhees *et al*, 2003). The universal route for glucose degradation was at one time considered to be the Embden–Meyerhof (EM) pathway (Fig. 1.02). In this pathway, glucose is phosphorylated to glucose-6-phosphate by an ATP dependent glucokinase. Phosphoglucose isomerase then converts glucose-6-phosphate into fructose-6-phosphate, which is phosphorylated by a second ATP dependent kinase to form fructose-1,6-bisphosphate. This is then cleaved into glyceraldehyde-3-phosphate and dihydroxyacetone phosphate, which is subsequently isomerised into a second molecule of glyceraldehyde-3-phosphate. Glyceraldehyde-3-phosphate is converted into pyruvate through the action of glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase, with a net yield of two mol ATP per mol glucose.

In 1952 an alternative pathway for glucose metabolism was discovered (Entner & Doudoroff, 1952) which was distinct from the classical EM pathway. Versions of this Entner-Doudoroff pathway have since been shown to exist in all domains of life (Danson *et al*, 2007). The distinction between these pathways is in the conversion of glucose to glyceraldehyde-3-phosphate as both pathways convert glyceraldehyde-3-phosphate into pyruvate, utilising the same enzymes.

In the classical Entner-Doudoroff (ED) pathway, glucose is phosphorylated by an ATP dependent glucokinase to glucose-6-phosphate, and this is then oxidised to 6-phosphogluconate by glucose-6-phosphate dehydrogenase. 6-Phosphogluconate dehydratase converts 6-phosphogluconate into 2-keto-3-deoxy-(6-phospho)-gluconate (KDPG), which is then cleaved into glyceraldehyde-3-phosphate and pyruvate by KDPG-aldolase and results in a net yield of one mol ATP per mol glucose.

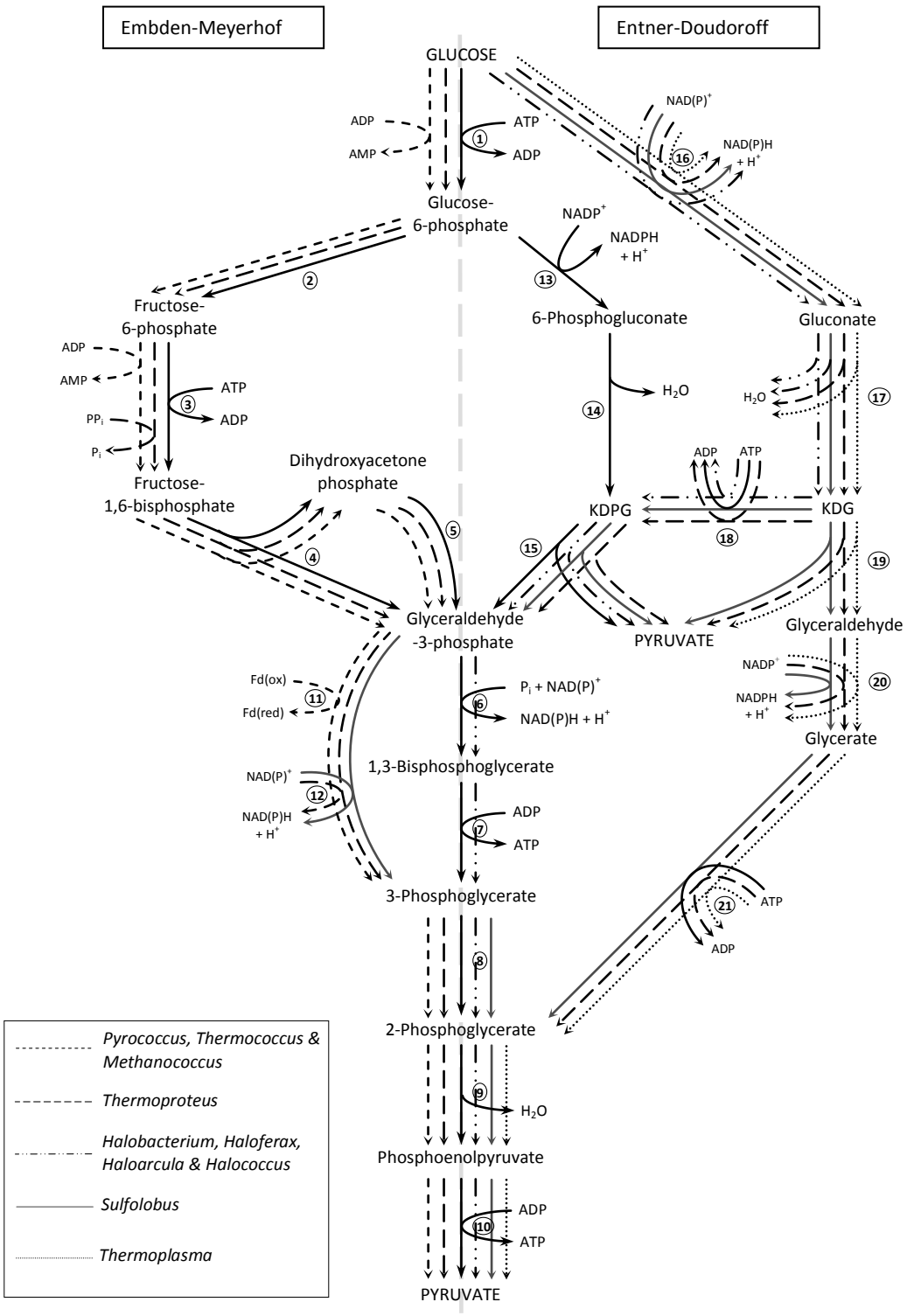


Fig. 1.02. Routes of glucose catabolism in the Archaea. The classical EM and ED pathways are shown in bold arrows, with modified pathways identified in individual organisms shown as dashed lines (taken from Danson et al, 2007). Numbered enzymes are: 1) glucokinase, 2) phosphoglucose isomerase, 3) phosphofructokinase, 4) fructose 1,6-bisphosphate aldolase, 5) triose-phosphate isomerase, 6) glyceraldehyde-3-phosphate dehydrogenase, 7) phosphoglycerate kinase, 8) phosphoglycerate mutase, 9) enolase, 10) pyruvate kinase, 11) glyceraldehyde-3-phosphate ferridoxin oxidoreductase, 12) nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase, 13) glucose-6-phosphate dehydrogenase, 14) 6-phosphogluconate dehydratase, 15) KDPG aldolase, 16) glucose dehydrogenase, 17) gluconate dehydratase, 18) KDG kinase, 19) KDG aldolase, 20) glyceraldehyde dehydrogenase, 21) glycerate kinase.

Modified versions of the EM and ED pathways have been identified in both aerobic and anaerobic archaea and, in at least one hyperthermophilic archaeon, have been shown to exist in parallel (Ahmed *et al*, 2004). While most of the intermediates in these modified pathways are the same as those found in the classical routes, the majority of archaeal enzymes involved in these reactions show no similarity to the corresponding 'classical' enzymes in bacteria and eukarya (Siebers & Schonheit, 2005), suggesting an independent evolution of these different catabolic routes (Verheers *et al*, 2003). The pyruvate formed by these various glycolytic reactions is oxidised to acetyl-CoA by pyruvate-ferredoxin oxidoreductase, which then enters into the citric acid cycle in oxygen, nitrate and sulphur reducing archaea (Siebers & Schonheit, 2005).

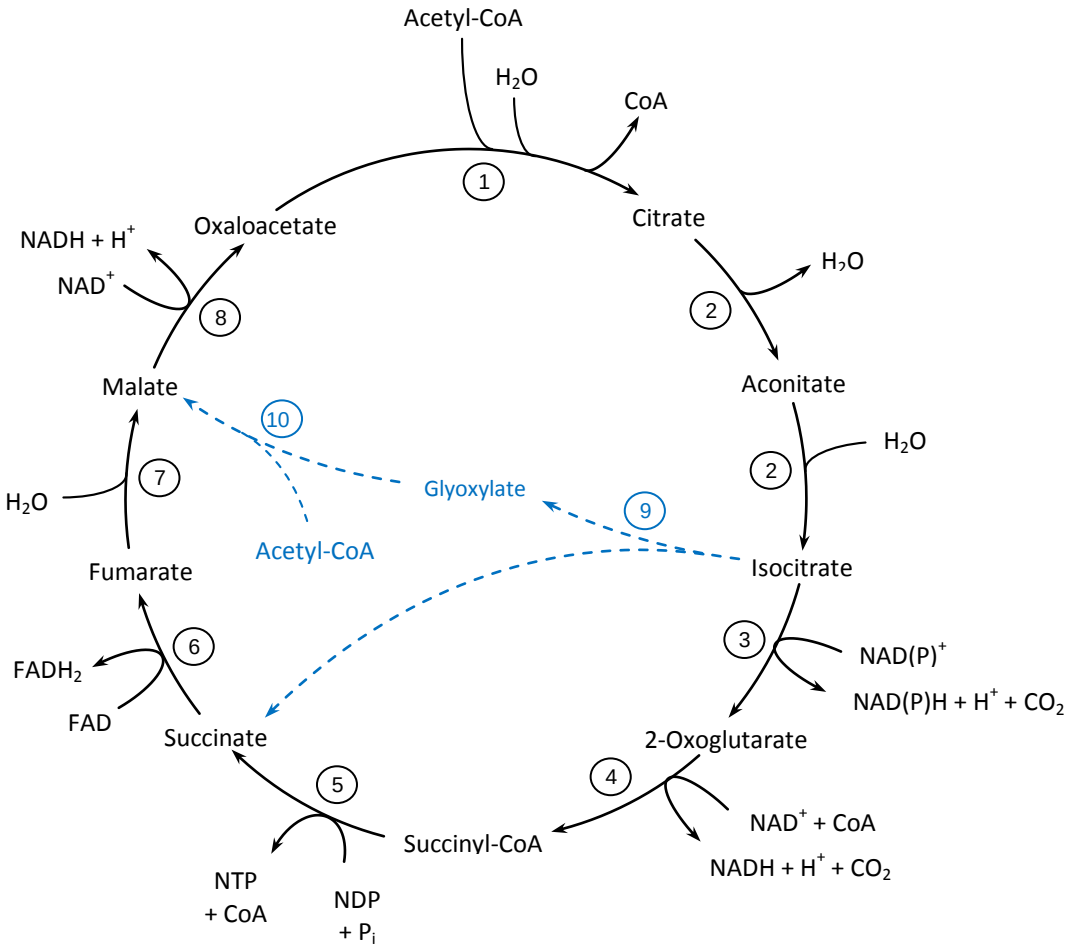


Fig. 1.03. The citric acid cycle and glyoxylate shunt (shown in blue). Enzymes in the citric acid cycle are: [1] Citrate synthase, [2] Aconitase, [3] Isocitrate dehydrogenase, [4] 2-Oxoglutarate dehydrogenase complex, [5] Succinate thiokinase, [6] Succinate dehydrogenase, [7] Fumarase, [8] Malate dehydrogenase. Enzymes involved in the glyoxylate shunt are [9] Isocitrate lyase and [10] Malate synthase.

The citric acid cycle (Fig. 1.03) was discovered in aerobic organisms and allows the complete oxidation of nutrients to CO_2 and H_2O with a higher yield of ATP than can be generated by either the EM pathway, or the various ED pathways present in archaea.

The citric acid cycle comprises a series of intermediates which undergo chemical transformations to allow the loss of 2 carbons (as CO_2) and the removal of hydrogen by $NAD(P)^+$ and FAD (the reoxidation of which with molecular oxygen yields ATP and H_2O), concluding in the regeneration of oxaloacetate. The glyoxylate shunt is a modification of the citric acid cycle creating a pathway which bypasses the steps involving the loss of CO_2. The existence of the glyoxylate cycle is well established in bacteria, but it has also been found to operate in some archaea including *Haloferax volcanii* (Serrano *et al*, 1998) and *Sulfolobus acidocaldarius* (Uhrigshardt *et al*, 2002).

Research to date has mainly focussed on catabolic routes for glucose in the archaea, while the pathways for the degradation of the naturally occurring C5 sugars D-xylose and L-arabinose remain largely unknown. In many bacteria, such as *Escherichia coli, Bacillus,* and *Lactobacillus* species, xylose is converted to xylulose 5-phosphate by xylose isomerase and xylulose kinase. This intermediate is further degraded by the pentose phosphate cycle or phosphoketolase pathway. Interestingly, xylulose 5-phosphate is also an intermediate of the most common L-arabinose degradation pathway in bacteria.

Recently, a metabolic pathway of D-xylose degradation was reported in the halophilic archaeon, *Haloferax volcanii.* In this pathway, D-xylose is converted to D-xylonate via D-xylonolactone by the sequential action of xylose dehydrogenase and xylonolactonase. Xylonate dehydratase then catalyses the formation of 2-keto-3-deoxy-xylonate (2-keto-3-deoxy-D-pentulosonic acid), which is dehydrated further to form 2-oxoglutarate-semialdehyde. 2-Oxoglutarate semialdehyde dehydrogenase then generates 2-oxoglutarate, an intermediate of the citric acid cycle. This is similar to the pathway for L-arabinose reported in species of *Azospirillum* where the metabolic enzymes involved are evolutionally related to equivalent enzymes in phosphorylative and non-phosphorylative versions of the Entner-Doudoroff pathway, indicating it may have evolved from different ancestral metabolic pathways (Watanabe *et al*, 2006).

1.3. *Sulfolobus solfataricus*

Sulfolobus solfataricus is a hyperthermophilic archaeon from the kingdom Crenarchaeota (Fig. 1.01). It is a strict aerobe that grows optimally at 80-85°C and between pH 2-4, but is able to maintain an intracellular pH of around 6.5 by generating a proton gradient across the cytoplasmic membrane (Moll & Schäfer, 1988). It has been reported to utilise numerous carbon sources including peptone, tryptone and yeast extract (Grogan, 1989), as well as carbohydrate substrates including the four most-commonly occurring sugars in nature, D-glucose, D-galactose, D-xylose and L-arabinose and also the non-natural isomer D-arabinose (Brouns *et al*, 2006).

In recent years, *S. solfataricus* has become one of the best studied members of the crenarchaeota due to its relatively straightforward growth requirements (Grogan, 1989), the availability of a complete genome sequence (She *et al*, 2001) and the development of genetic tools that permit the homologous expression of genes (Albers *et al*, 2006). This has resulted in *Sulfolobus* becoming the model organism for research on mechanisms for DNA replication, transcription and translation (She *et al*, 2001), as well as the study of archaeal metabolism (Lamble *et al*, 2005; Brouns *et al,* 2006, Zaparty *et al,* 2010).

1.4. Sugar metabolism in *S. solfataricus*

Aerobic archaea are known to use modified versions of the Entner-Doudoroff pathway for carbohydrate metabolism (Ahmed *et al*, 2005), including semi-phosphorylative and non-phosphorylative variants. In contrast to both the classical and semi-phosphorylative pathways, which produce one mol of ATP per mol of glucose, the non-phosphorylative pathway results in no net gain of ATP.

Early studies on *S. solfataricus* showed that glucose was metabolised via the non-phosphorylative Entner-Doudoroff pathway (De Rosa, 1984) and subsequent studies suggest that a similar pathway operates in *Sulfolobus acidocaldarius* (Selig, 1997). In this pathway (Fig. 1.04) glucose dehydrogenase (GDH) catalyses the $NAD(P)^+$ dependent oxidation of glucose to gluconate, which is then converted to 2-keto-3-deoxy-gluconate (KDG) by the action of gluconate dehydratase (GD). KDG-aldolase cleaves KDG into pyruvate and glyceraldehyde, the latter of which is converted to glycerate by glyceraldehyde oxidoreductase. Glycerate kinase utilises a molecule of ATP to achieve the phosphorylation of glycerate to 2-phosphoglycerate, which is subsequently dehydrated to

phosphoenolpyruvate by the action of an enolase. Finally, pyruvate kinase converts phosphoenolpyruvate to pyruvate, synthesising one mol of ATP.

More recent evidence suggests that *Sulfolobus* operates a split non-phosphorylative pathway as both KDG and 2-keto-3-deoxygalactonate (KDGal) can be phosphorylated by a single kinase to 6P-KDG and 6P-KDGal, and that the KDG-aldolase will also catalyse the cleavage of these two metabolites to pyruvate and glyceraldehyde-3P, although the regulation of these pathways remains to be elucidated (Lamble *et al*, 2005; Danson *et al*, 2007).

In contrast to other organisms in which metabolic enzymes are generally inducible and highly substrate specific, enzymes from the modified ED pathway in *S. solfataricus* are constitutively expressed (Lamble *et al*, 2003). Kinetic analyses of GDH, GD and KDG-aldolase also showed these enzymes are capable of catalysing the catabolism of galactose, the C4 epimer of glucose, to pyruvate and glyceraldehyde. These findings have lead to the suggestion that the non-phosphorylative ED pathway exhibits a metabolic promiscuity towards these two hexose sugars (Fig. 1.04).

Studies on the recombinantly produced GDH show that it also has high activity with the pentose sugars D-xylose and L-arabinose and the recombinant KDG-aldolase has been found to catalyse the condensation of the C-2 compound glycolaldehyde and pyruvate to form 2-keto-3-deoxypentulosonic acid [H.J. Lamble, D.W. Hough, S.D. Bull and M.J. Danson, unpublished data]. These findings imply both enzymes are likely to have a role in the catabolism of the naturally occurring pentose sugars, D-xylose and L-arabinose, as well as of the hexose sugars D-glucose and D-galactose. Research on the purified native GD however, revealed that it was specific for gluconate and galactonate.

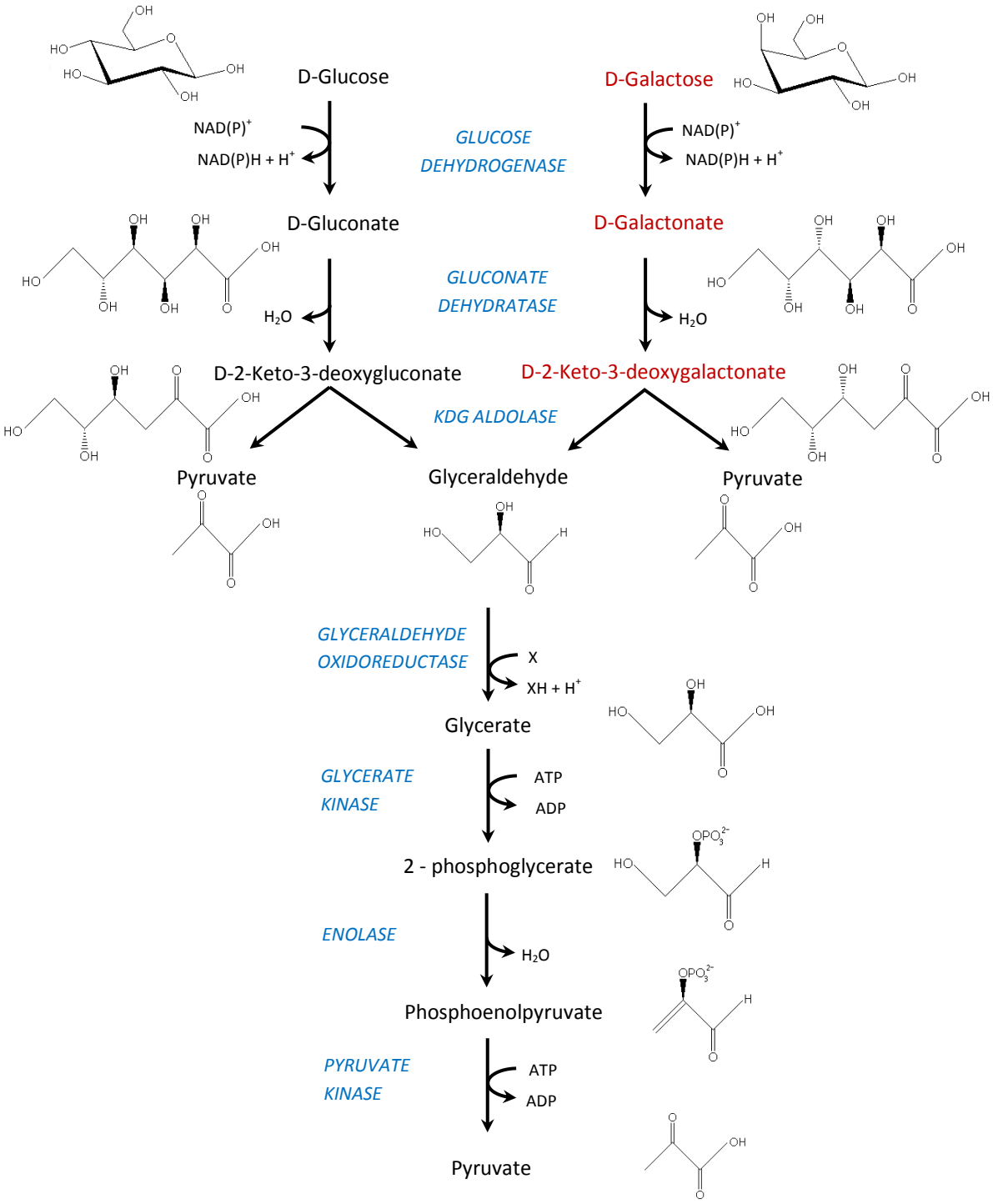


Fig. 1.04. The non-phosphorylative Entner-Doudoroff pathway in *S. solfataricus* for the catabolism of glucose and galactose (Modified from Lamble *et al*, 2003). Glucose dehydrogenase catalyses the $NAD(P)^+$ dependent oxidation of glucose to gluconate or galactose to galactonate, which are then converted to 2-keto-3-deoxy-gluconate (KDG) or 2-keto-3-deoxy-galactonate (KDGal) by the action of gluconate dehydratase (GD). KDG-aldolase cleaves KDG or KDGal into pyruvate and glyceraldehyde, the latter of which is converted to glycerate by glyceraldehyde oxidoreductase. Glycerate kinase utilises a molecule of ATP to achieve the phosphorylation of glycerate to 2-phosphoglycerate, which is subsequently dehydrated to phosphoenolpyruvate by the action of an enolase. Finally, pyruvate kinase converts phosphoenolpyruvate to pyruvate, synthesising a molecule of ATP.

1.5. Proposal for a new pathway of C5 metabolism

Catabolism of the non-natural isomer D-arabinose via an inducible series of enzymes has also been reported in *S. solfataricus* (Brouns *et al*, 2006). In this pathway, D-arabinose is dehydrogenated by a specific D-arabinose dehydrogenase to D-arabinonate; this is dehydrated by a specific D-arabinonate dehydratase and the resulting 2-keto-3-deoxy-D-pentulosonic acid is converted to 2-oxoglutarate through the action of a second dehydratase and a 2,5-dioxopentanoate dehydrogenase (DOPDH). However, the complete catabolic route of the naturally occurring sugars D-xylose and L-arabinose remains unclear.

The promiscuous activities reported for glucose dehydrogenase and KDG-aldolase (Lamble *et al*, 2003) suggest that there may be an alternative metabolic pathway for C5 sugars similar to that reported in the bacterium *Rhizobium japonicum* (Pedrosa & Zancan, 1974). The presence of a D-xylonate / L-arabinonate dehydratase in *Sulfolobus* would complete the first part of the pathway by producing 2-keto-3-deoxy-D-pentulosonic acid and 2-keto-3-deoxy-L-pentulosonic acid, which would both by cleaved by KDG-aldolase into pyruvate and glycolaldehyde. While glyceraldehyde catabolism from the C6 route is well established, the fate of glycolaldehyde remains unknown. Based on known enzymes from other organisms, a new pathway has been proposed for the metabolic fate of glycolaldehyde (Fig. 1.05).

In this new pathway, glycolaldehyde would be oxidised to glycolate by glycolaldehyde oxidoreductase, glycolate would then be converted to glyoxylate by glycolate dehydrogenase and finally malate synthase would condense acetyl-CoA and glyoxylate to form malate. Malate would enter the citric acid cycle, for complete oxidation to CO_2, NADH and $FADH_2$ (Fig. 1.06).

This pathway would provide a catabolic route for both naturally occurring C5 sugars D-xylose and L-arabinose, leading directly into central metabolism.

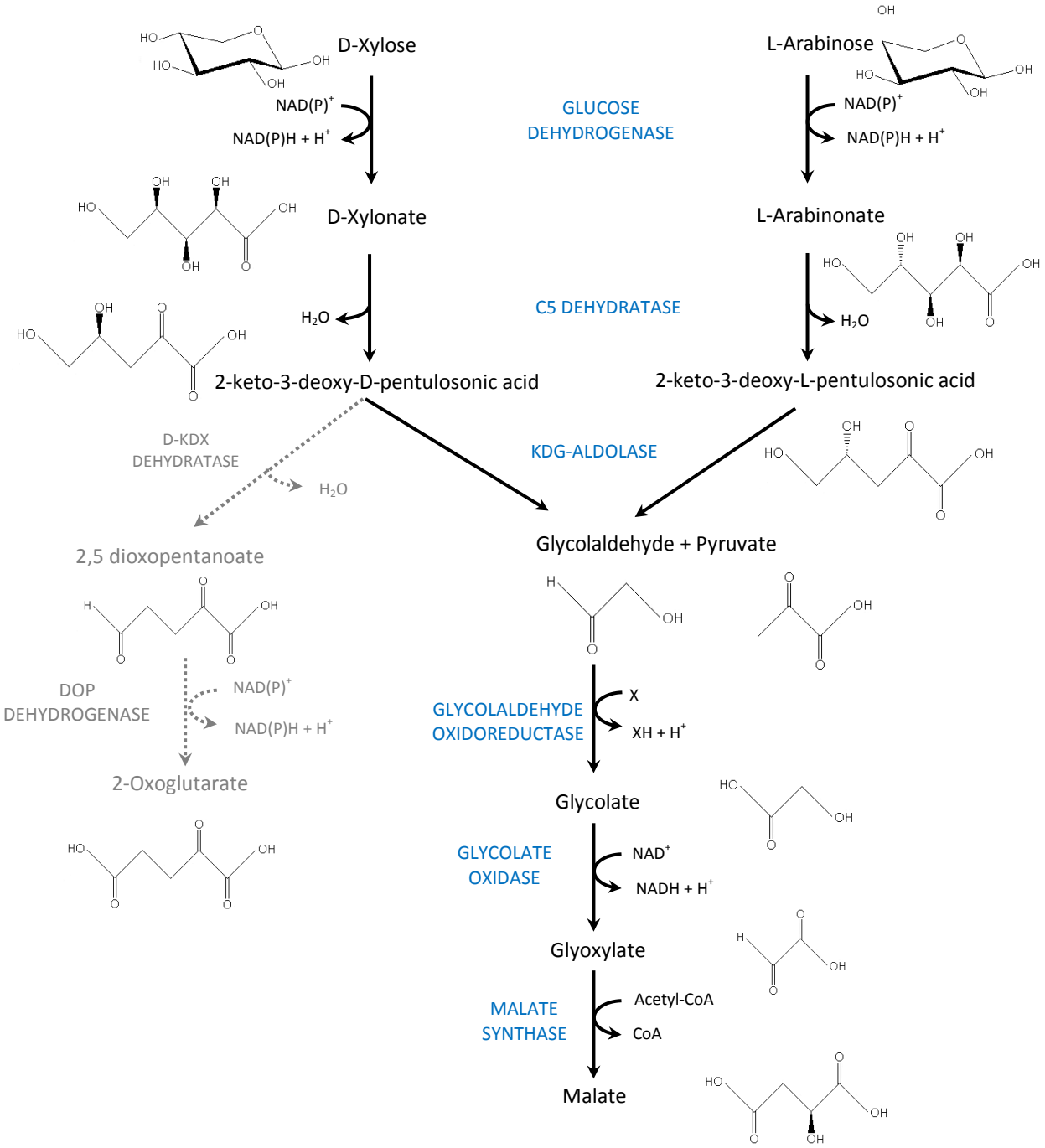


Fig. 1.05. Routes of C5 catabolism in *S. solfataricus.* The route proposed by Brouns et al (2006) is shown in grey. Proposal for a new pathway of C5 catabolism is shown in black. The promiscuous glucose dehydrogenase would generate D-xylonate and L-arabinonate from D-xylose and L-arabinose. A D-xylonate / L-arabinonate dehydratase would then dehydrate these products to form either the D or L enantiomer of 2-keto-3-deoxypentulosonic acid which KGD-aldolase would convert into glycolaldehyde and pyruvate. Glycolaldehyde is then oxidised to glycolate by glycolaldehyde oxidoreductase, glycolate would then be converted to glyoxylate by glycolate dehydrogenase and finally malate synthase would condense acetyl-CoA and glyoxylate to form malate, which would enter the citric acid cycle for complete oxidation to CO_2, NADH and $FADH_2$ (Fig. 1.06).

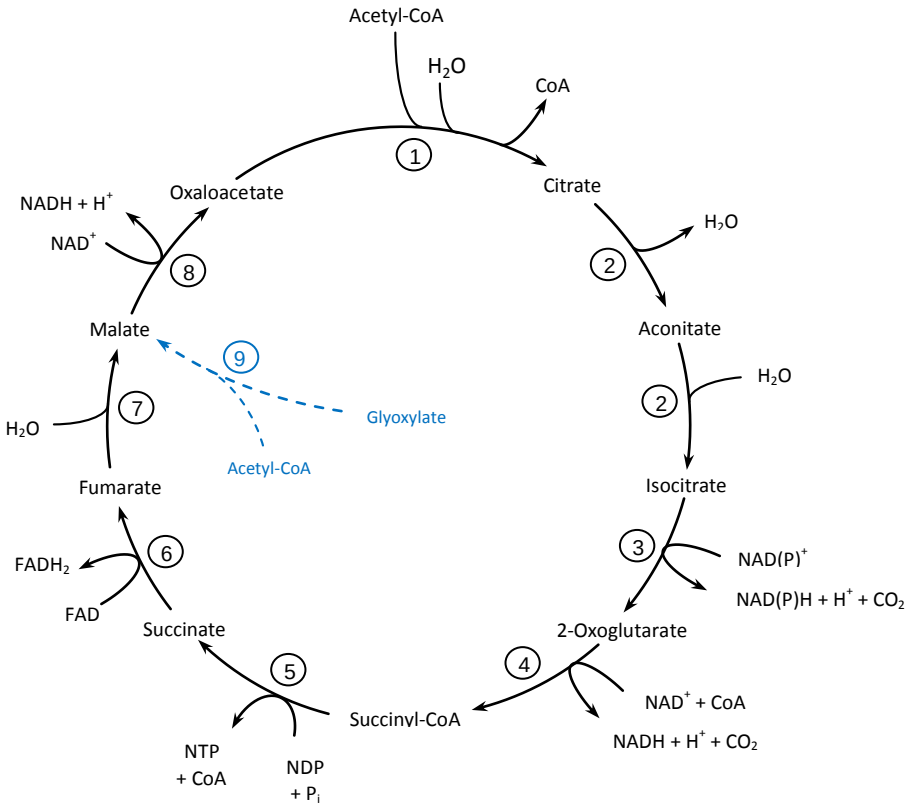


Fig. 1.06. Glyoxylate generated by C5 metabolism in *S. solfataricus* is condensed to malate by malate synthase, and then enters into the citric acid cycle for complete oxidation to CO_2, NADH and $FADH_2$. Enzymes involved are numbered; 1) Citrate synthase, 2) Aconitase, 3) Isocitrate dehydrogenase, 4) 2-Oxoglutarate DH complex, 5) Succinate thiokinase, 6) Succinate dehydrogenase, 7) Fumarase, 8) Malate dehydrogenase, 9) Malate synthase.

1.6. Extremozymes

The survival of extremophiles requires that their cellular machinery, from transcription / translation systems to metabolic enzymes, operates at extreme temperatures. Extremophiles (from both archaea and bacteria) have therefore become a major source of novel enzymes for use in biocatalysis and molecular biology, due the ability of these 'extremozymes' to function under diverse conditions (e.g. temperature, salt concentration) which may be beneficial for industrial processes (van den Burg, 2003). In industry, the use of thermophilic enzymes permits the use of higher process temperatures, which may have favourable effects on the rate of reaction, the solubility of the substrate (e.g. starches) and has the additional benefit of lowering the risk of microbial contamination (Eijsink *et al*, 2004).

Perhaps the best known example of extremozymes for a molecular biology application is that of *Taq* DNA polymerase, isolated from the thermophilic bacterium *Thermus aquaticus*. However, it is not just high temperature functions that are sought after; chaperonins identified from psychrophilic bacteria are used in protein expression to enhance solubility and folding when co-expressed in an *E. coli* host grown at 12°C (Ferrer *et al*, 2003).

The identification of structural adaptations of these enzymes that allow them to function optimally at extreme temperatures is an ongoing and active topic of research. The comparison of mesophilic and thermophilic enzymes has often been used as a source to identify features that could account for this enhanced stability, although there does not seem to be a single unique feature that allows thermophilic enzymes to remain stable at temperatures up to 100°C (Unsworth *et al*, 2007). Instead, several factors such as conformational flexibility, increased hydrophobicity and increased numbers of ionic interactions all seem to contribute towards enhancing stability (Hough & Danson, 1999).

1.7. Implications of enzyme promiscuity

The specificity of enzymes was once considered to be at the very foundation of catalysis, but the increasing number of examples of promiscuous enzymes reported in recent years has lead to the acceptance that a large proportion of enzymes will show secondary activity or alternative catalytic mechanisms towards other substrates (e.g. O'Brian & Herschlag, 1999; Lamble *et al*, 2003; Aharoni *et al*, 2005).

Enzyme promiscuity is likely to confer a selective advantage to organisms in changing environments, enabling their survival and continued evolution. The gene duplication and subsequent natural mutation of these promiscuous enzymes can cause deviations of function that enhance genetic diversity and subsequently drive the evolution of the divergent new enzyme (Jensen, 1976). The initial presence of promiscuity has therefore been associated with evolution of enzyme function (Aharoni *et al*, 2005).

The existence of enzyme superfamilies supports the idea of divergent evolution (Gerlt, 2004). Enzymes belonging to a superfamily will quite often have low sequence similarity and catalyse different overall reactions but share structural and functional features (O'Brien & Herschlag, 1999). Members of the enolase superfamily, for example, catalyse different overall reactions through a variety of mechanisms but all share a partial reaction

in which an active site base abstracts the α-proton of the carboxylate substrate to generate an enolate anion intermediate that is stabilised by coordination to an essential Mg^{2+} ion (Gerlt, 2004).

As well as conferring a selective advantage to organisms, promiscuous enzymes are often used as a source to develop new enzymatic activities through mutagenesis (Serrano *et al*, 1993; Bloom & Arnold, 2009) for industrial applications, although often at the cost of enzyme stability (Tokuriki *et al*, 2008). A greater understanding of promiscuity is therefore essential to increase understanding of molecular recognition, the biochemistry of living systems and the process of evolution (Nobeli *et al*, 2009).

1.8. Aims of project

Through the numerous metabolic studies carried out on *S. solfataricus,* the routes of C6 sugar metabolism within this organism have been established. To date there is, however, only limited information available relating to the routes of pentose metabolism in *Sulfolobus*, especially for the naturally occurring sugars D-xylose and L-arabinose.

The main aim of this project is therefore to elucidate the pathway for D-xylose and L-arabinose catabolism in *S. solfataricus* based on the route proposed earlier in this chapter. This catabolic route will be investigated in the first instance by detecting the presence of enzyme activities in the native organism. The second part of this project will be concerned with the characterisation of recombinantly produced enzymes involved in the pathway, both for comparison with metabolic enzymes in other organisms and to explore further enzymes that may potentially be interesting as biocatalysts.

Finally, using rational design and site-directed mutagenesis, the promiscuity and thermoactivity of selected recombinant enzymes will be explored to try to increase the understanding of the relationship between promiscuity, activity and thermostability in thermophilic enzymes.

Chapter 2: General Materials and Methods

All chemicals were supplied by Sigma-Aldrich (Gillingham, UK) unless otherwise stated.

2.1. Measurement of enzyme activities

Xylose dehydrogenase activity was determined as described by Lamble *et al* (2003). Assays were performed at 70°C in 1ml of 100mM HEPES (Melford, Suffolk, UK) pH8.0. The assay mixture contained 0.5mM $NADP^+$ and 5mM D-xylose, and the reaction was started by the addition of 10µl enzyme. The reduction of $NADP^+$ was followed by the increase in absorbance at 340nm (ε = 6220 $M^{-1}cm^{-1}$). One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

Xylonate dehydratase activity was determined using the method of Lamble *et al* (2004). Assays were performed in 400µl of 100mM Bis-Tris, pH7.0 containing 10mM $MgCl_2$ and 5mM calcium D-xylonate (Acros Organics, Geel, Belgium) or calcium L-arabinonate (Pfaltz & Bauer, Connecticut, USA). The reaction was started by the addition of 10µl of enzyme and incubated for 10 min at 70°C before being stopped by transferring 100µl of the reaction mixture to 10µl of 12% (w/v) trichloroacetic acid and centrifuged at 16000 x g for 5 min. The formation of 2-keto-3-deoxy-D/L-pentulosonic acid was quantified spectrophotometrically at 549nm (ε = 67800 $M^{-1}cm^{-1}$) by the reaction with thiobarbituric acid, as described by Buchanan *et al* (1999). One unit of enzyme activity is defined as the amount of enzyme required to form of 1µmol of 2-keto-3-deoxypentulosonic acid per min.

KDG-aldolase activity was determined using the method of Buchanan *et al* (1999). Assays were performed in 400µl of 100mM sodium phosphate buffer, pH6.0 containing 37mM glycolaldehyde (Fluka, Gillingham, UK) and 37mM sodium pyruvate. The reaction was started by the addition of 10µl of enzyme and incubated for 10 min at 70°C before being stopped by transferring 100µl of the reaction mixture to 10µl of 12% (w/v) trichloroacetic acid and centrifuged at 16000 x g for 5 min. The formation of 2-keto-3-deoxy-D/L-pentulosonic acid was quantified spectrophotometrically at 549nm (ε = 67800 $M^{-1}cm^{-1}$) by the reaction with thiobarbituric acid. Determination of kinetic parameters was performed between 0.5-37mM glycolaldehyde in the presence of 37mM pyruvate or 0.5-37mM pyruvate in the presence of 37mM glycoaldehyde. One unit of enzyme activity is defined

as the amount of enzyme required to form of 1µmol of 2-keto-3-deoxypentulosonic acid per min.

Aldehyde oxidoreductase was assayed at 70°C in 1 ml of 100mM Bis-Tris buffer, pH7.0, 2mM 2,6-dichlorophenolindophenol (DCPIP) and 5mM glycolaldehyde (Fluka, Wisconsin, US). Glycolaldehyde was allowed to equilibrate at temperature for 2 min before the addition of DCPIP to record the background rate. Enzyme was added after additional 1 min incubation, and the reduction of the DCPIP was followed at 595nm (ε = 22000 M^{-1} cm^{-1}). Determination of kinetic parameters was performed between 0.1-10mM glycolaldehyde in the presence of 50µM DCPIP. One unit of enzyme activity is defined as the amount of enzyme required to reduce 1µmol of dichlorophenolindophenol per minute.

2,5-Dioxopentanoate dehydrogenase (2-oxoglutarate semialdehyde dehydrogenase) assays were performed at 70°C in 1ml of 100mM HEPES, pH7.5, as described by Brouns *et al* (2006). The assay mixture contained 1mM $NADP^+$ and 10mM pentanedial, and the reaction was started by the addition of 10-20µl enzyme. The reduction of $NADP^+$ was followed by the increase in absorbance at 340nm. One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

Glyoxylate reductase activity was determined as described by Yoshikawa *et al* (2007). Assays were performed at 70°C in 1ml of 100mM HEPES, pH7.5. The assay mixture contained 0.4mM NADH and 10mM glyoxylate, the reaction was started by the addition of 10µl enzyme. The oxidation of NADH was followed by the decrease in absorbance at 340nm (ε = 6220 $M^{-1}cm^{-1}$). Determination of kinetic parameters was performed between 0-10mM glyoxylate in the presence of 0.4mM NADH or from 0-0.4mM NADH in the presence of 50mM glyoxylate. One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

Malate synthase activity was determined as described by Uhrigshardt *et al* (2002). The assay was performed at 70°C in 1ml buffer containing 20mM EPPS, 100mM KCl, pH8.0. The assay mixture contained 10mM DTNB, 0.14mM acetyl-CoA and 10mM glyoxylate, the reaction was started by the addition of 10µl enzyme and followed by the increase in absorbance at 412nm (ε = 13600 $M^{-1}cm^{-1}$). Determination of kinetic parameters was performed between 0.01-10mM glyoxylate in the presence of 140µM acetyl-CoA or from 2-

140µM acetyl-CoA in the presence of 10mM glyoxylate. One unit of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

Isocitrate lyase activity was determined as described by Uhrigshardt *et al* (2002). The assay was performed at 70°C in 1ml of 100mM Bis-Tris, pH7.5 containing 100mM NaCl and 20mM $MgCl_2$. The assay mixture contained 2.5mM phenylhydrazine, 15mM isocitrate and the reaction was started by the addition of 10µl enzyme sample. The reaction was followed by the increase in absorbance at 324nm (ε = 17000 $M^{-1}cm^{-1}$) corresponding to the production of glyoxylatephenylhydrazone. Determination of kinetic parameters was performed between 0.1-15mM isocitrate in the presence of 2.5mM phenylhydrazine. One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of glyoxylatephenylhydrazone per minute.

2.2. Thermoactivity and thermostability assays

Thermoactivity was determined using the standard enzyme assays detailed above performed over a range of temperatures between 50-95°C. The cuvette containing the buffer was equilibrated to the desired temperature before the addition of substrates. These were again allowed to reach the desired temperature (measured with a Whatman DT100 thermocoupler) before the addition of 10-25µg of enzyme. This was briefly mixed and initial rates were obtained by calculating the absorbance change per min over the first 20s of the reaction.

Thermostability was measured by incubating 10-25µg of pure enzyme at 85°C or 95°C in a 0.5ml thin-walled PCR tube (Biohit, Devon, UK) for 1-180 min before immediately transferring to ice to prevent further denaturation. Precipitated protein was removed by centrifugation at 13683 x g for 5 min and the supernatant assayed at 70°C using the standard assay detailed above.

2.3. SDS PAGE

A 10% running gel was prepared by the addition of 10% (v/v) acrylamide, 325mM Tris buffer (pH8.8), 33.1% (v/v) MilliQ water, 0.1% (w/v) SDS, 0.1% (v/v) TEMED (National Diagnostics, Georgia, USA) and 0.2% (w/v) ammonium persulphate. A 4% Stacking gel was

prepared by the addition of 2.25% (v/v) acrylamide, 78mM Tris buffer (pH6.8), 45% (v/v) MilliQ water, 0.06% (w/v) SDS, 0.05% (v/v) TEMED and 0.1% (w/v) ammonium persulphate.

Samples were prepared by heating at 100°C for 1.5 min with 2x SDS-PAGE buffer containing 50mM Tris, 4% SDS, 20% (w/v) Sucrose, 10% (v/v) β-mercaptoethanol (BDH Laboratory Supplies) and 0.2% (w/v) Bromophenol Blue (BDH Prolabo, Leicestershire, UK). Gels were run at 35mA per gel, in 25mM Tris buffer containing 0.25M glycine (Fisher Scientific, Leicestershire, UK) and 0.1% (w/v) SDS.

Protein bands were visualised by first staining the gel in Coomassie Blue stain containing 0.25% (w/v) Coomassie Brilliant Blue R-250, 45% (v/v) methanol, 45% (v/v) water and 10% (v/v) acetic acid and then destaining in a solution containing 70% (v/v) water, 20% (v/v) methanol and 10% (v/v) acetic acid.

2.4. Protein concentration estimation

Protein concentrations were determined by the Bradford Method (Bradford, 1976), using reagent supplied by Bio-Rad Laboratories (Hercules, CA, USA). Bovine serum albumin (Pierce, IL, USA) was used as a standard diluted to 100µl/ml from a 2mg/ml stock to create a standard curve of 0-100µl/ml using a 2:7 dilution of Bradford Reagent. Dilutions of protein samples were prepared in distilled H_2O. The absorbance of standards and samples was read at 595nm (Cary 50 Bio, Varian).

2.5. Determination of native molecular weights by gel filtration

All products were supplied by GE Healthcare (Bucks, UK) unless otherwise stated.

500µl of sample was loaded onto a Superdex 200 10/300GL (10mm x 300mm) gel filtration column using 50mM Tris / HCl (pH8.5) and a flow rate of 0.4ml / min. The void volume was determined by loading 500µl of 1mg/ml Dextran Blue and determined to be 7.7ml. The column was calibrated with low and high molecular weight standards shown below (table 2.01) according to the manufacturer's instructions.

Protein	M_r
Carbonic anhydrase	29000
Albumin	66000
Alcohol dehydrogenase	150000
Amylase	200000
Apoferritin	443000
Thyroglobulin	669000
Blue Dextran 2000	>2000000

Table 2.01. Relative molecular mass of protein standards used for the calibration of Superdex 200 10/300GL (10mm x 300mm) gel filtration column

The K_{av} of standards were calculated using the equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

Where V_o = column void volume, V_e = elution volume and V_c = geometric column volume and the values plotted against the Log M_r of each standard to form a calibration curve (Fig 2.01).

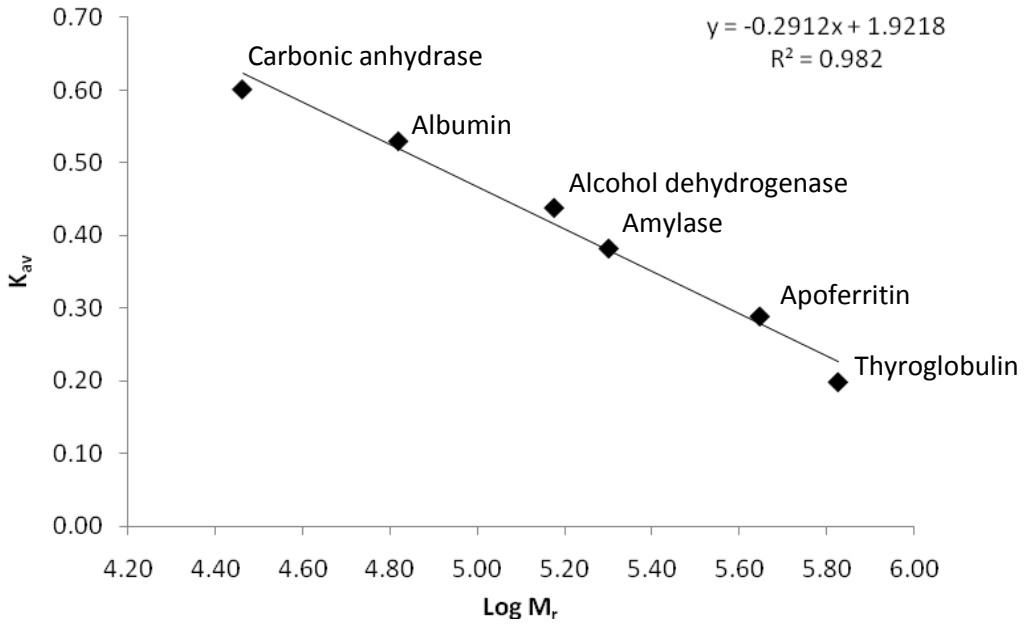


Fig. 2.01. Calibration curve for Superdex 200 10/300GL (10mm x 300mm) gel filtration column using protein standards described in table 2.01.

A Log M_r value for each sample was estimated using the equation of the calibration curve in the formula:

$$\text{Log } M_r = \frac{(V_e - 1.9218)}{0.2912}$$

Chapter 3: Demonstrating the complete C5 Entner-Doudoroff pathway at the level of enzyme activities

3.1. Introduction

As previously discussed in Chapter 1, the pathways of glucose catabolism in *S. solfataricus* are well established and have highlighted interesting and unique enzymes with respect to their apparent substrate promiscuity (Lamble *et al,* 2003; Milburn *et al*, 2006). Despite a recent study proposing a catabolic route for the non natural sugar D-arabinose (Brouns *et al*, 2006), the complete pathway for D-xylose and L-arabinose metabolism has not yet been elucidated.

There are a growing number of studies developing computation models to describe changes in metabolism or protein expression levels in response to environmental fluctuations. Perhaps most notable and relevant to this work is that of the *Sulfolobus* Systems Biology project (SulfoSYS) undertaken by Albers *et al* (2009), which seeks to produce an *in silico* cell model of central carbohydrate metabolism in *Sulfolobus* and its variation with temperature (Albers *et al*, 2009; Zaparty *et al,* 2010). However, the descriptive potential of these models is proportional to the knowledge of the components and pathways being modelled (Downs, 2006). It is therefore vital that, like the data presented here, protein function and metabolic interactions continue to be a focus of biochemical research to enhance the understanding of these metabolic networks.

3.2. Materials and Methods

All chemicals were supplied by Sigma-Aldrich (Gillingham, UK) unless otherwise stated.

3.2.1. Growth of *S. solfataricus* and preparation of cell extracts

Freeze-dried *S. solfataricus* P2 (DSM 1617) was obtained from DSMZ (Germany) and grown in basal salts medium (Brock *et al*, 1972) containing (per litre) 0.28g KH_2PO_4, 1.30g $(NH_4)_2SO_4$, 0.25g $MgSO_4.7H_2O$, 0.25g $CaCl_2.7H_2O$, and adjusted to pH3.5. The medium also contained a 1% (w/v) trace element solution (20.0mg $FeCl_3.6H_2O$, 4.5mg $Na_2B_4O_7.10H_2O$, 1.8mg $MnCl_2.4H_2O$, 0.22mg $ZnSO_4.7H_2O$, 0.05mg $CuCl_2.2H_2O$, 0.03mg $Na_2MoO_4.4H_2O$, 0.03mg $VOSO_4.2H_2O$, 0.01mg $CoSO_4.7H_2O$), 0.05% (w/v) Yeast extract and 2% (w/v) D-glucose, D-xylose or L-arabinose. Cells were harvested by centrifugation at 5000 x g for 10 min and resuspended at 0.2g/ml in 50mM Tris-HCl, pH8.0 and incubated for 30 min. Cells were lysed by four 30-s bursts of sonication using a 150-watt Ultrasonic Disintegrator (MSE Scientific Instruments, Crawley, UK) and a cell extract obtained by centrifugation at 12520 x g for 30 min.

3.2.2. Phosphatase treatment of C5 dehydratase

40µl of partially purified enzyme was incubated at 37°C in 50mM Tris-HCl containing 100mM NaCl, 10mM $MgCl_2$ and 1mM DTT for 0-40min in the presence of 0.4U calf intestinal phosphatase (CIP) (NEB, Massachusetts, USA). The negative control was incubated under the same conditions without the addition of the CIP.

3.3. Results

3.3.1. Enzyme activities

Investigating the proposed pathway of C5-sugar catabolism in *S. solfataricus* began with determining the native enzyme activities in cell extract. *S. solfataricus* was grown on D-glucose, D-xylose or L-arabinose and cell extracts prepared as described in Materials and Methods. Enzyme activities were measured in these unfractionated extracts and are shown in Fig. 3.01. The identification and kinetic analysis for each of these enzymes are described below.

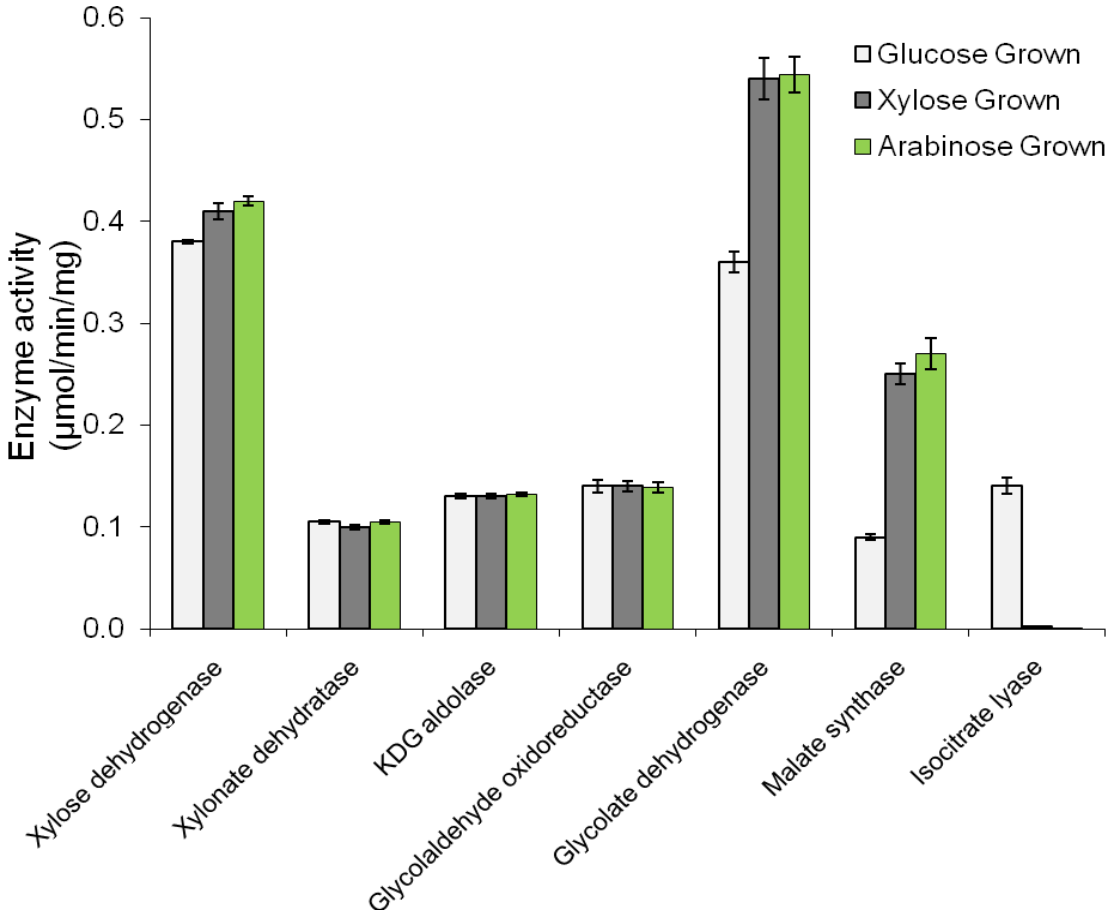


Fig. 3.01. Enzyme activities determined from a cell extract of *S. solfataricus* grown on either C6 or C5 sugars as a carbon source (as described in Materials and Methods). Errors are shown as the difference between duplicates of the activity assay on the same cell extract.

3.3.2. Xylose Dehydrogenase

Xylose dehydrogenase activity was detected in a cell extract of *S .solfataricus* grown on C5 sugars as described in General Materials and Methods. The K_M for xylose with $NADP^+$ as a cofactor was found to be 0.18 (±0.005)mM, corresponding to previously reported data (Lamble *et al* 2003). As described in the Introduction, a glucose dehydrogenase (gene I.D: SSO3003) with equally high activity with D-glucose, D-galactose, D-xylose and L-arabinose has previously been cloned and the recombinant protein characterised (Lamble *et al* 2003). The crystal structure of the enzyme complexed with D-glucose and D-xylose has also been determined to resolutions of 1.6 and 1.5 Å, respectively, and the structural basis of its substrate promiscuity elucidated (Milburn *et al*, 2006). No further investigation has therefore been undertaken of this enzyme in relation to the C5 pathway proposed here.

3.3.3. Xylonate dehydratase

A gluconate dehydratase enzyme (gene I.D: SSO3198*)* has previously been purified and identified from *Sulfolobus solfataricus* (Lamble *et al*, 2004). The substrate promiscuity displayed by the glucose dehydrogenase and KDG-aldolase from the same pathway led to the assumption that a single promiscuous enzyme catalysed the dehydration of gluconate to 2-keto-3-deoxygluconate and xylonate to 2-keto-3-deoxyxylonate (2-keto-3-deoxy-D-pentulosonic acid). The purified native dehydratase enzyme was fully characterised and although activity was shown with both gluconate and galactonate, no activity could be detected with D-xylonate or L-arabinonate.

Using the assay described in chapter 2, C5 dehydratase activity was detected in a cell extract of *S. solfataricus* grown on C5 sugars with D-xylonate or L-arabinonate as substrates (Fig. 3.01). Partial purification of a *Sulfolobus solfataricus* cell extract revealed that this xylonate dehydratase is distinct from the gluconate dehydratase previously characterised by Lamble e*t al* (2004). The C6 and C5 dehydratase activities can be separated by both anion exchange (Fig. 3.02) and gel filtration, which give M_r values of 48,000 for the gluconate dehydratase and 130,000 for xylonate dehydratase.

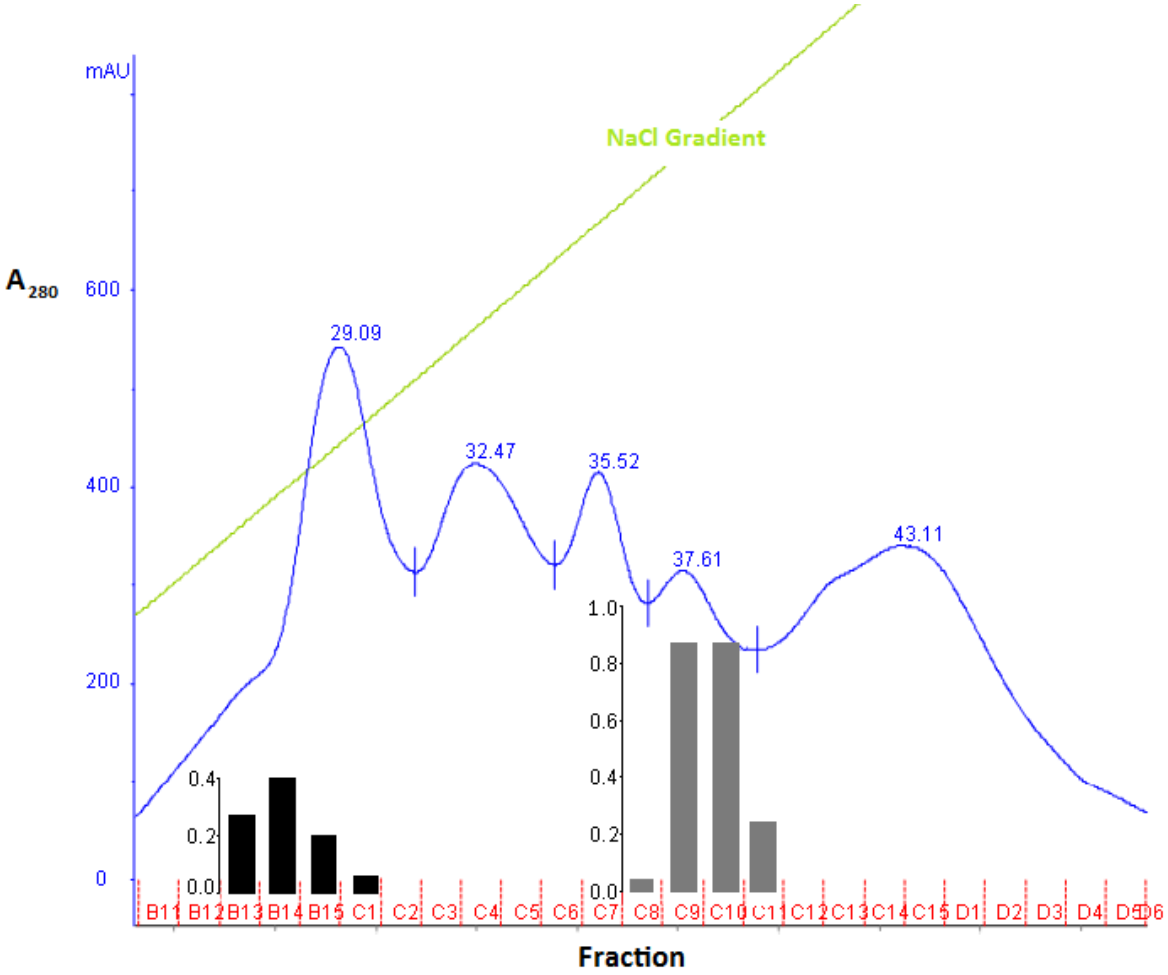


Fig. 3.02. Anion exchange profile for the partial purification of a cell extract from *S. solfataricus* grown on C5 sugars. The blue traces with corresponding retention volumes (ml) above represents change in absorbance at 280nm with an increasing NaCl gradient (shown as a green line). The 1ml fractions containing the separated proteins were all assayed for both xylonate and gluconate dehydratase activity. The superimposed histograms represent the activity (U/ml) found in each fraction for xylonate dehydratase (black) and gluconate dehydratase (grey).

The partially purified C5 dehydratase showed activity with both D-xylonate and L-arabinonate, but no activity with D-gluconate or D-galactonate. Activity for this enzyme is determined by a stopped assay (as described in chapter 2). It is, therefore, important to ensure that the assay is recording the initial rate of reaction so that kinetic parameters can be calculated using the Michaelis-Menten equation.

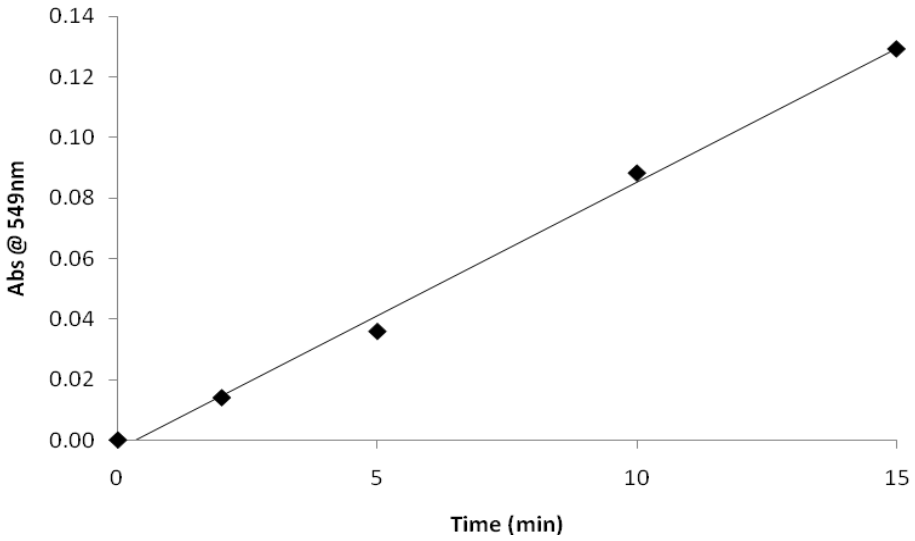


Fig. 3.03. Absorbance change over the first 15 minutes of the xylonate dehydratase assay with 5mM xylonate as a substrate at 70°C as described in Materials and Methods.

Fig. 3.03 shows that the increase in absorbance (proportional to the production of 2-keto-3-deoxypentulosonic acid) remains linear over the duration of the assay (10min) under standard assay conditions, allowing data from single time-point measurements to be used to calculate initial rates.

Due to an apparent decrease in enzyme velocity at higher substrate concentrations (Fig. 3.04), which was not due to a pH change over the duration of the assay, the data were analysed according to the substrate inhibition equation:

$$v = \frac{V_{\max}[S]}{K_M + [S] + \frac{[S]^2}{K_{si}}}$$

(Equation 1)

where K_{si} is the dissociation constant for ES_2, the substrate-inhibited enzyme. The following kinetic parameters were obtained: D-xylonate K_M = 0.93 (± 0.47) mM, K_{si} = 4.8 (± 2.4) mM; L-arabinonate K_M = 0.27 (± 0.10), K_{si} = 14.5 (± 6.2) mM.

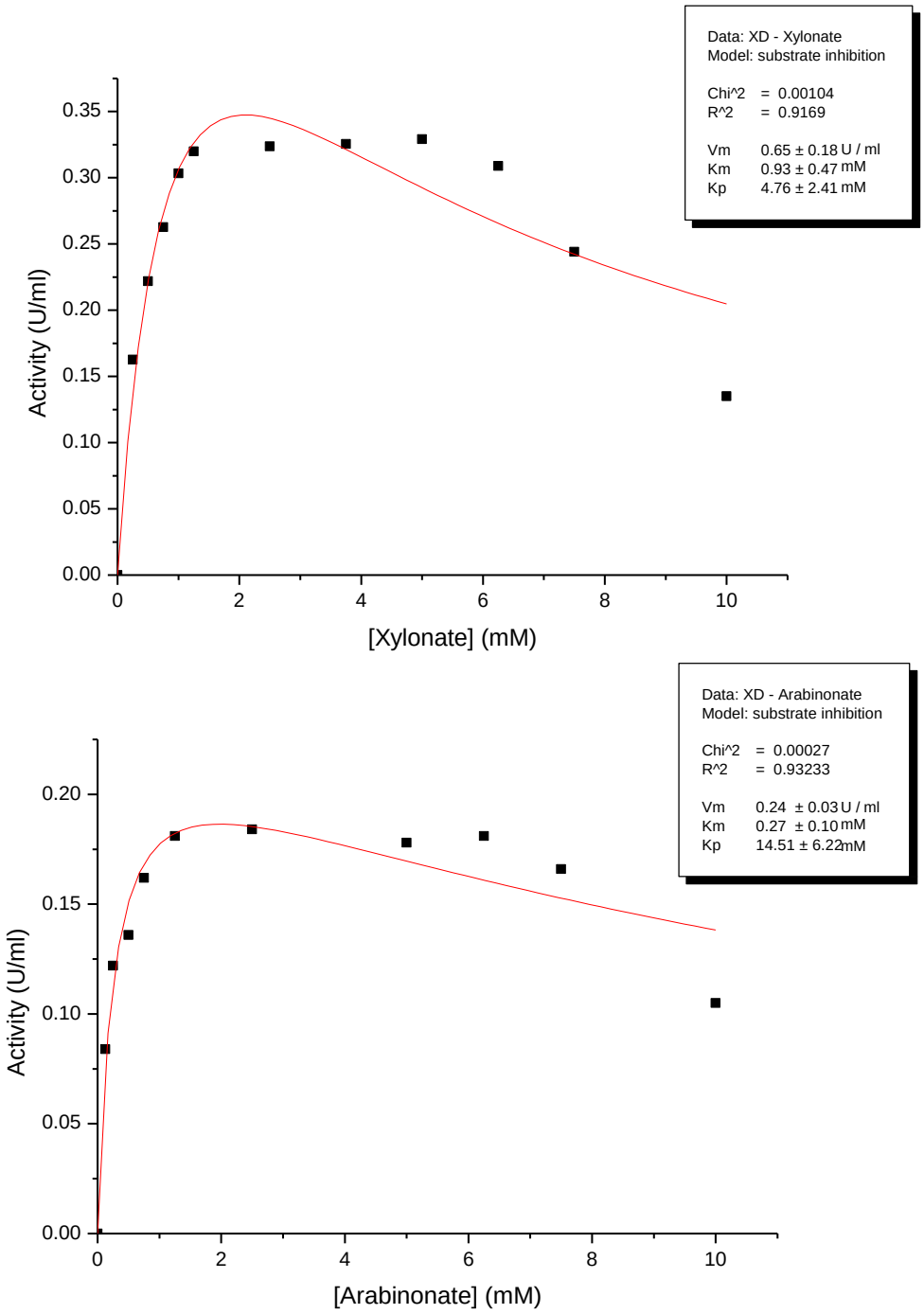


Fig. 3.04. Activity of C5 dehydratase with xylonate and arabinonate with increasing substrate concentration, fitted to the substrate inhibition equation (p. 21). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce of 1μmol of 2-keto-3-deoxypentulosonic acid per minute.

It is clear from the plots in Fig. 3.04 and the calculated errors that the data do not give a good fit to the classical substrate inhibition equation and so true values for K_M and V_{max} could not be calculated using this method.

Studies on the gluconate dehydratase (Kim & Lee, 2005) have suggested a possible phosphorylation / dephosphorylation system controlling the activity of the enzyme. This was tested with the semi-purified C5 dehydratase by incubating with calf intestinal phosphatase (CIP) for 0 - 40 min at 37°C (as described in Materials and Methods), and then assaying the remaining activity after this time. The C5 dehydratase was incubated under the same conditions in the absence of CIP to serve as a control for the reaction (Fig. 3.05).

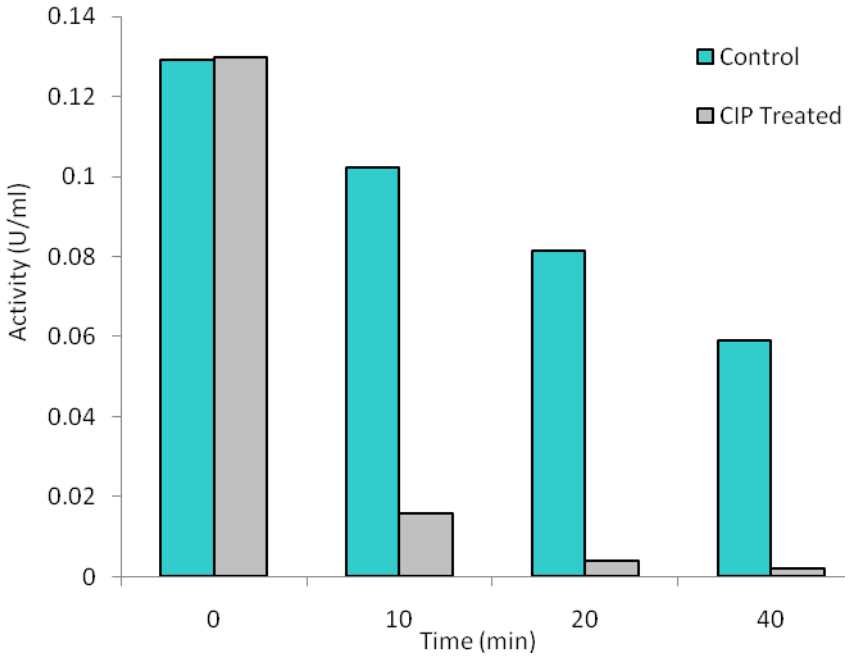


Fig. 3.05. Remaining dehydratase activity after incubation at 37°C either with or without phosphatase (CIP) added. Activity was detected as described in Materials and Methods with 4.5mM D-xylonate. One unit (U) of enzyme activity is defined as the amount of enzyme required to produce of 1µmol of 2-keto-3-deoxypentulosonic acid per minute.

The unstable nature of the semi-purified C5 dehydratase can clearly be seen by the 50% reduction of activity observed over a period of 40 min in the absence of CIP. However, the effect of CIP can still clearly be seen by the 88% reduction in dehydratase activity after only 10 minutes (compared to the control at 10min), indicating that the enzymes does require phosphorylation for activity.

Several attempts were made to purify the dehydratase further for identification, but due to the rapid loss of activity even in the presence of phosphatase inhibitors it was not possible to achieve the required enrichment or purity to allow conclusive identification.

The known gluconate dehydratase protein sequence was used in a homology based searches of the *S. solfataricus* genome, and a candidate gene for the C5 dehydratase has been identified (SSO2665) which when translated shows a 62% identity at the amino acid level with the gluconate dehydratase.

3.3.4. KDG-Aldolase

A KDG-aldolase gene (SSO3197) has previously been cloned and expressed, and the recombinant protein characterised (Lamble *et al* 2003) and shown to have activity with both glyceraldehyde and glycolaldehyde. Using the assay described in chapter 2, KDG-aldolase activity was detected in a cell extract of *S. solfataricus* grown on C5 sugars with glycolaldehyde and pyruvate as substrates (table 3.01 and Fig. 3.01).

The following data were produced using the recombinant KDG-aldolase, purified to homogeneity. Activity for this enzyme is determined by a stopped assay (as described in chapter 2). It is, therefore, important to ensure that the assay is recording the initial rate of reaction so that kinetic parameters can be calculated using the Michaelis-Menten equation. Fig. 3.06 shows that the increase in absorbance (proportional to the production of 2-keto-3-deoxypentulosonic acid) remains linear over the duration of the assay (10min) under standard assay conditions, allowing data from single time-point measurements to be used to calculate initial rates.

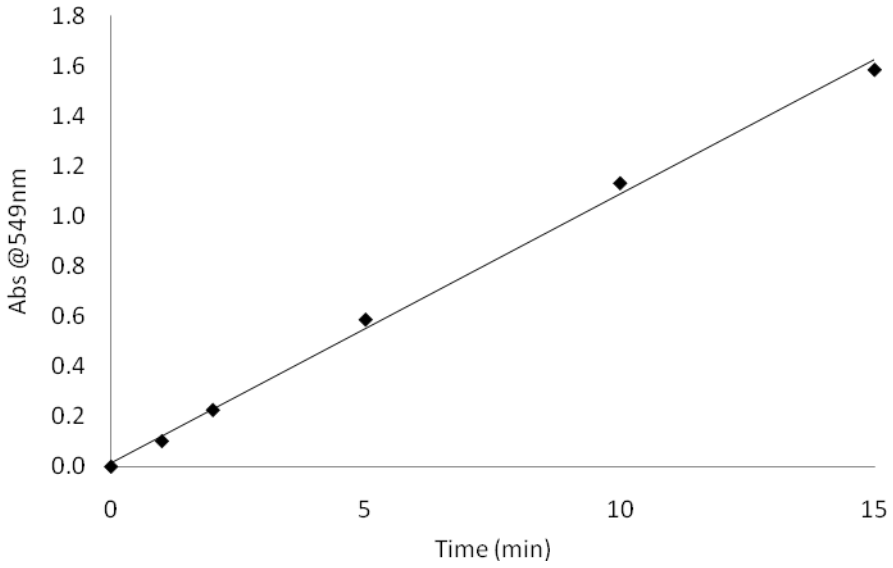


Fig. 3.06. Absorbance change over the first 15 minutes of the recombinant KGD-aldolase assay at fixed concentrations of glycolaldehyde (15mM) and pyruvate (37mM).

Enzyme velocity appeared to decrease at high concentrations of glycolaldehyde, but not at high concentrations of pyruvate. The kinetic data for glycolaldehyde were analysed according to Equation 1 (p. 21) but were not a good fit to the substrate inhibition model (Fig. 3.08) producing values with errors over 50%.

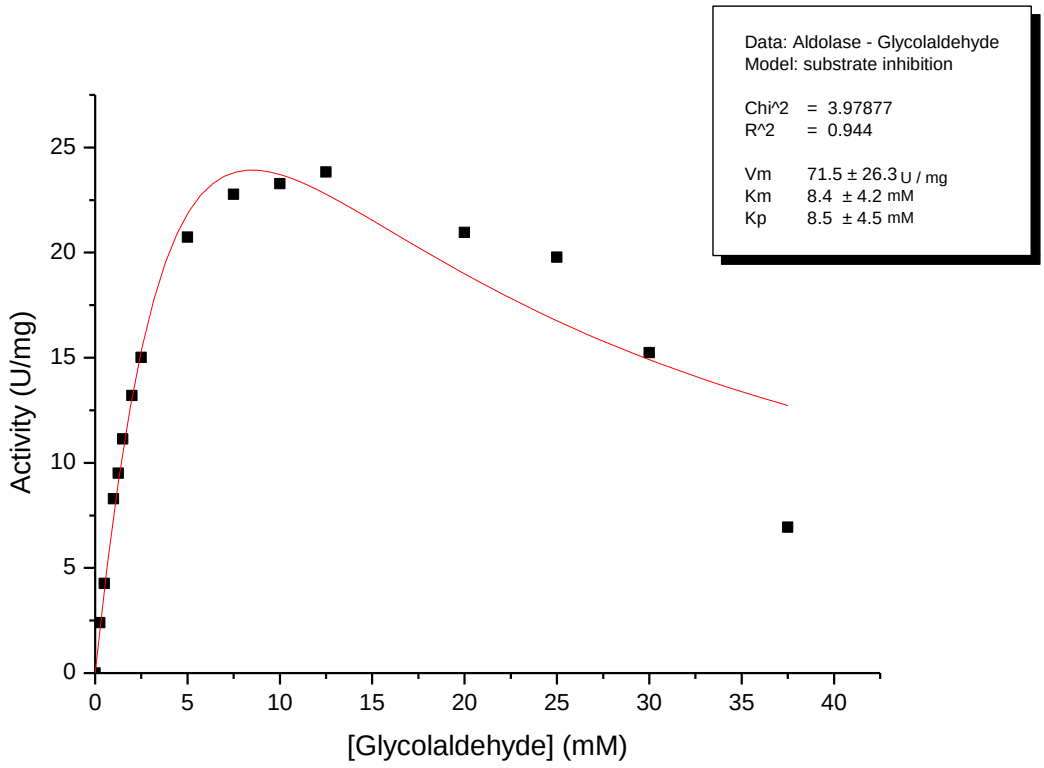


Fig. 3.08. Activity of KDG-aldolase with increasing glycolaldehyde concentration, fitted to the substrate inhibition equation (p. 21). One unit of enzyme activity is defined as the amount of enzyme required to form of 1µmol of 2-keto-3-deoxypentulosonic acid per minute.

The following kinetic parameters were calculated; K_M = 8.4 (± 4.2) mM, K_{si} = 8.5 (± 4.5) mM, but due to the poor fit, a true K_M could not be determined using this model. This inhibition also affected determination of true kinetic parameters for pyruvate, as glycolaldehyde could not be present in excess. Instead, apparent values were calculated at a lower concentration of glycolaldehyde (10mM) (Table 3.02), resulting in an underestimation of V_{max}.

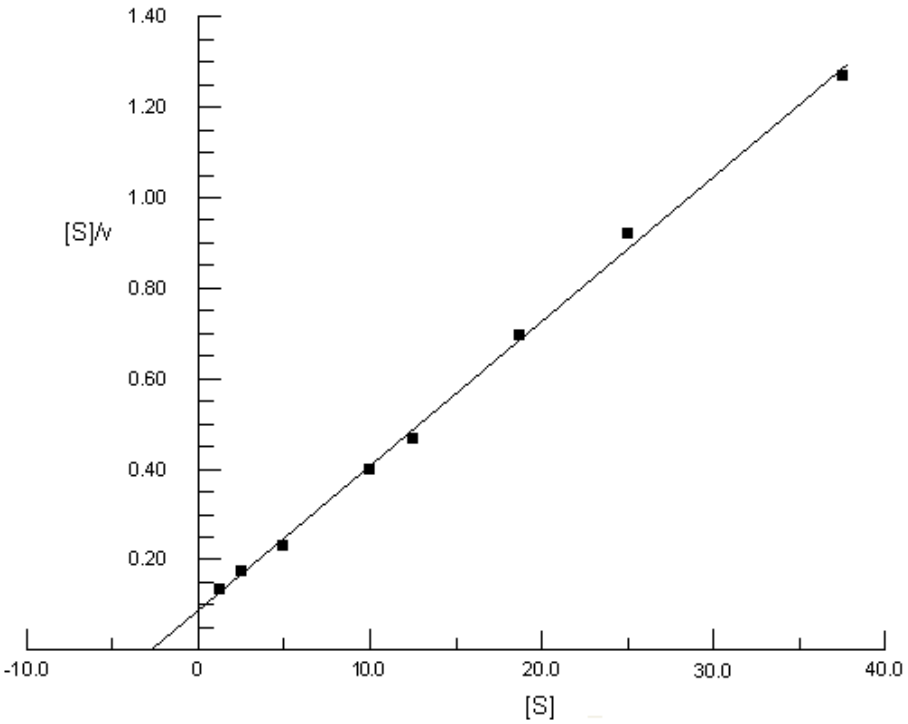


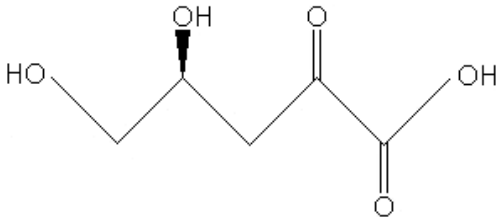
Fig. 3.09. Hanes - Woolf plot for the recombinant KGD-aldolase with 0-37mM pyruvate and 10mM glycolaldehyde at 70°C as described in Materials and Methods. [S] Is shown in mM and v in U/ml where one unit of enzyme activity is defined as the amount of enzyme required to form of 1μmol of 2-keto-3-deoxypentulosonic acid per minute.

Substrate	K_M^{APP} **(mM)**	V_{max}^{APP} **(U/mg)**
Glycolaldehyde	8.4 (± 4.2)	71.5 (± 26.3)
Pyruvate	2.9 (± 0.2)	31.5 (± 0.4)

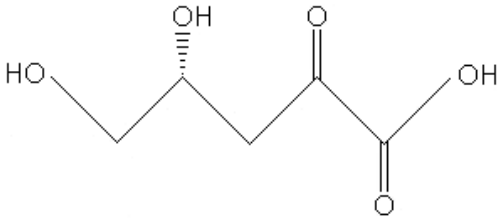
Table 3.02. Apparent kinetic parameters for the recombinant KGD-aldolase with glycolaldehyde (calculated using the substrate inhibition equation in Origin® software) and pyruvate, (determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit of enzyme activity is defined as the amount of enzyme required to form of 1μmol of 2-keto-3-deoxypentulosonic acid per minute.

3.3.5. Coupled assay of KDG-aldolase with xylonate dehydratase

The C5 substrates 2-keto-3-deoxy-D-pentulosonic acid and 2-Keto-3-deoxy-L-pentulosonic acid are not commercially available and so it was not possible to assay the enzyme directly in the forwards direction. It is assumed that both D and L 2-keto-3-deoxypentulosonic acid are produced by the condensation of glycolaldehyde and pyruvate, but as these compounds are enantiomers (Fig. 3.10) they are not easily distinguishable as they display the same chemical properties.



2-keto-3-deoxy-D-pentulosonic acid



2-Keto-3-deoxy-L-pentulosonic acid

Fig. 3.10. Structure of the two C5 compounds utilised by KDG-aldolase, 2-keto-3-deoxy-D-pentulosonic acid and 2-keto-3-deoxy-L-pentulosonic acid, which are enantiomers due to the single chiral centre.

To confirm that the enzyme was able to catalyse the cleavage of both substrates, the semi-purified native xylonate dehydratase was utilised in a coupled assay. The dehydratase reaction was allowed to proceed under standard conditions with either D-xylonate or L-arabinonate for a period of 10 minutes. At this point a sample was taken and the reaction stopped. Pure KDG-aldolase was added to the remainder of the assay and the reaction was allowed to continue for a further 5 minutes before the assay was stopped. The concentrations of 2-keto-3-deoxy-D-pentulosonic acid or 2-keto-3-deoxy-L-pentulosonic acid were then quantified in both assays.

To ensure that the observed decrease in product was due to the activity of the aldolase and not due to product degradation, a 15min dehydratase assay was repeated with each substrate. Aldolase activity was also measured in control reactions containing each of the substrates but no dehydratase enzyme (Table 3.03, Fig. 3.11).

	Concentration of 2-keto-3-deoxypentulosonic acid (μM)	
	D-xylonate substrate	**L-arabinonate substrate**
Control (no dehydratase)	0.0	0.0
10min dehydratase assay	4.0	1.1
15min dehydratase assay	4.7	1.3
10min dehydratase assay + 5min aldolase assay	0.4	0.1

Table 3.03. The concentration of 2-keto-3-deoxy-D-pentulosonic acid or 2-keto-3-deoxy-L-pentulosonic acid after a 10 min standard dehydratase assay with either D-xylonate or L-arabinonate as a substrate, and with the addition of KDG-aldolase for an additional 5 min. The equivalent product concentration after a 15 min dehydratase assay is also shown.

Addition of the aldolase to the dehydratase reaction resulted in a ten-fold decrease in product concentration in 5 minutes for both substrates, demonstrating that the aldolase can utilise both products of the dehydratase reaction.

The enzymatic rate of each reaction was also calculated (table 3.04). Evaluation of relative activities with the two substrates shows that that both the dehydratase and the aldolase have 30% of the activity with L-arabinonate / 2-keto-3-deoxy-L-pentulosonic acid compared with D-xylonate / 2-keto-3-deoxy-D-pentulosonic acid.

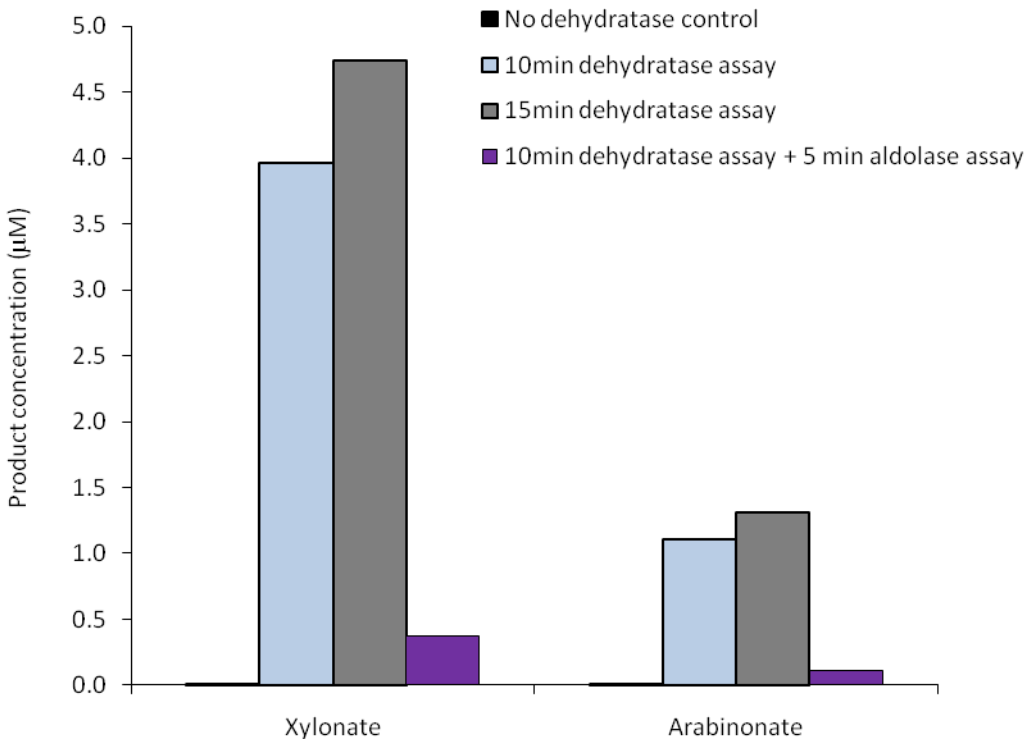


Fig. 3.11. The concentration of KDX or KDA after a 10min standard dehydratase assay with either D-xylonate or L-arabinonate as a substrate, the same assay with aldolase added for an additional 5min and the equivalent product concentration after a 15minute dehydratase assay.

	Activity (U/ml)	
	Xylonate	**Arabinonate**
Xylonate Dehydratase	0.25	0.07
Recombinant Aldolase	5.48	1.49

Table 3.04. Enzymatic rates for xylonate dehydratase and KDG-aldolase under standard assay conditions with either D-xylonate or L-arabinonate as a substrate.

The dramatic difference in activities shown in the table are due to the relative difference in concentration and purity of the two enzymes, the dehydratase being semi-purified from a *S. solfataricus* cell extract and the aldolase being a recombinantly-produced enzyme purified to homogeneity.

3.5.6. Glycolaldehyde Oxidoreductase

An annotation search of the *S. solfataricus* genome revealed only one annotated aldehyde dehydrogenase gene (SSO3117). This enzyme has already been characterised as the 2,5-dioxopentanoate (DOP) dehydrogenase in the alternative C5 pathway described by Brouns *et al* (2006) (chapter 1). The recombinant enzyme was shown to have secondary activities with other aldehydes including glyceraldehyde and glycolaldehyde, suggesting that it may in fact play a bi-functional role in C5 metabolism. This enzyme activity was detected in C5 grown *S. solfataricus* cell extract, using the method of Brouns *et al* (2006) and pentanedial as an artificial substrate for DOP (which is not commercially available). When the assay was repeated with glycolaldehyde as a substrate, activity was too low to be accurately measured. It is therefore unlikely that this enzyme is responsible for the conversion of glycolaldehyde to glycolate in this pathway.

However, Kardinahl *et al* (1999) reported the presence of a molybdenum containing aldehyde oxidoreductase in *Sulfolobus acidocaldarius* which was assayed with DCPIP as an artificial electron acceptor. The enzyme was purified and found to be composed of three subunits and the corresponding genes identified in the genome of *S. acidocaldarius* (Kardinahl *et al,* 1999).

Glycolaldehyde oxidoreductase activity was identified in *S. solfataricus* using DCPIP as an artificial electron acceptor and a blast search of the *S. solfataricus* genome revealed that three genes showing sequence identity to those in *S. acidocaldarius* were present (SSO2636, SSO2636, SSO2636). The *S. acidocaldarius* enzyme was reported to have a pH optimum of ~pH7 but the activity assay required some optimisation as glycolaldehyde exists in equilibrium between monomeric and dimer forms and spontaneously reduces DCPIP under a range of assay conditions. Although at the temperature of the assay (70°C) glycolaldehyde favours the monomeric form (Dr S Royer, University of Bath, personal communication), a range of buffer conditions was tested to minimise the background rate and allow accurate enzyme activity data to be collected (Table 3.05).

Assay condition	Background rate	Enzyme Rate	Adjusted rate	U/mg
pH 6 (50mM Phosphate)	0.23	0.16	-0.08	-0.03
pH 7 (50mM Phosphate)	0.26	0.19	-0.07	-0.02
pH 7 (100mM Bis-Tris)	0.08	0.48	0.39	0.14
pH 7 (100mM HEPES)	0.45	0.52	0.07	0.02
pH 8 (100mM HEPES)	0.55	0.61	0.06	0.02

Table 3.05. Optimisation of assay conditions for glycolaldehyde oxidoreductase at 70°C, using 2.5mM glycolaldehyde and 50µM DCPIP at 70°C, as described in general Materials and Methods. One unit (U) of enzyme activity is defined as the amount of enzyme required to reduce 1µmol of dichlorophenolindophenol per minute.

Partial fractionation of a C5 grown cell extract by gel filtration showed that the glycolaldehyde oxidoreductase activity corresponded to a protein of ~178KDa, very similar to that reported for *S. acidocaldarius*. This partially purified protein was used for the determination of kinetic parameters (Fig. 3.13). A decrease in enzyme velocity was apparent at high concentrations of glycolaldehyde. Analysis of these data using Equation 1 (pg 21) indicates that substrate inhibition could account for the decreased velocity observed at higher concentrations of glycolaldehyde (Fig. 3.13).

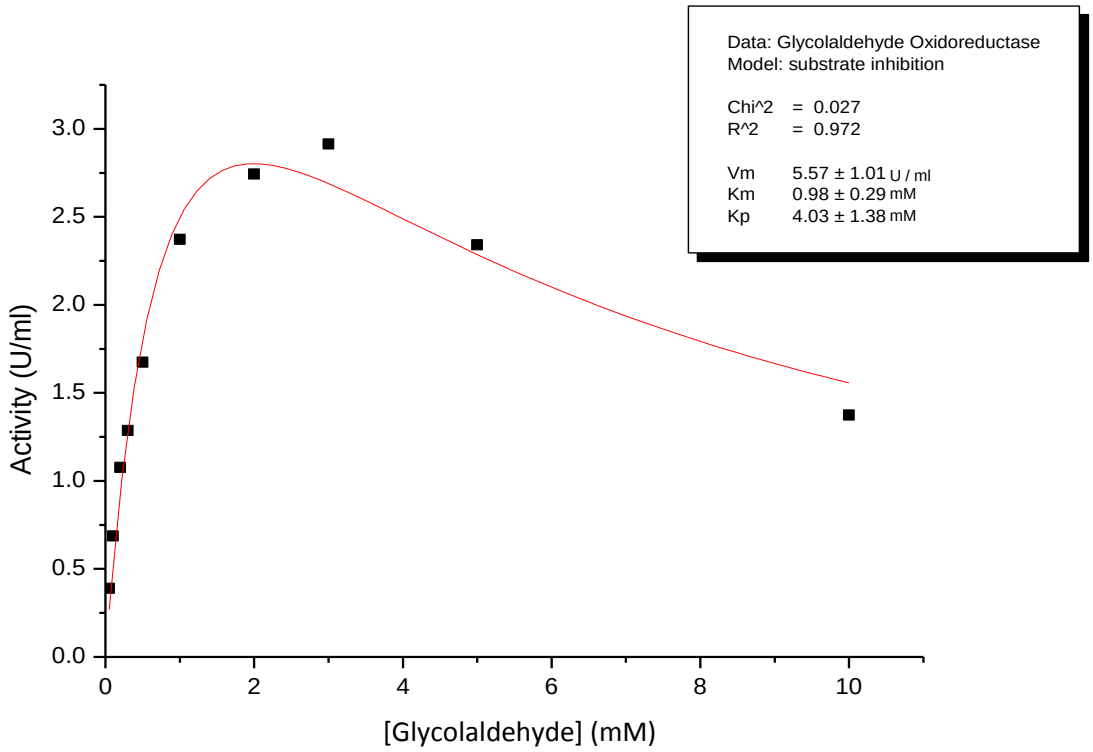


Fig. 3.13. Activity vs [S] plot for recombinant glycolaldehyde oxidoreductase. Activity was measured at 70°C in 100mM Bis-Tris pH7 at 0-10mM glycolaldehyde and 50µM DCPIP as described in general Materials and Methods. One unit (U) of enzyme activity is defined as the amount of enzyme required to reduce 1µmol of dichlorophenolindophenol per minute.

However, the calculated errors between 30% - 50% are too high to say conclusively that observed decrease in velocity is caused exclusively by substrate inhibition. Using the substrate inhibition model to calculate kinetic parameters; K_M = 0.98 (± 0.29) mM, V_{max} = 5.57 (±1.01) U/ml and K_{si} = 4.03 (± 1.38) mM.

In addition to the decrease in velocity, there was also a significant increase in the background rate of reaction at higher substrate concentrations (Fig. 3.15), which is likely to have affected the accuracy of calculating enzyme velocity under these conditions.

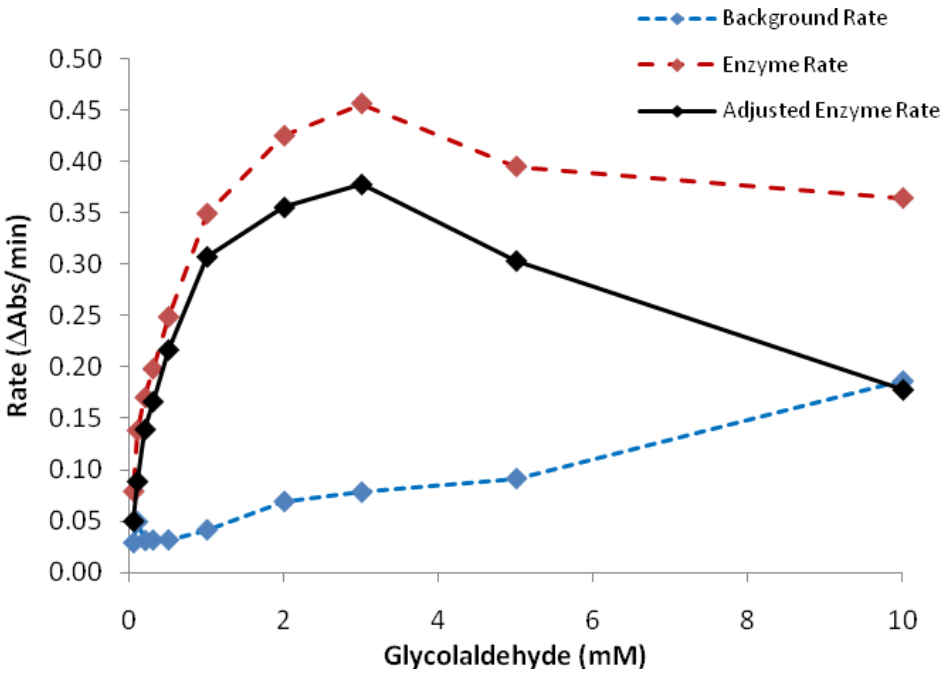


Fig. 3.15. Increase in enzymatic rate (red dashed) and background rate of reaction (blue dashed) with increase in substrate concentration using the standard assay described in Materials and Methods. The adjusted rates used to calculate enzyme activity are represented by the black line.

As reported in *S. acidocaldarius*, glycolaldehyde oxidoreductase was also found to have activity with glyceraldehyde (Fig. 3.16), although in the case of *S. solfataricus* enzyme, activity with glyceraldehyde appeared to be lower than that found with glycolaldehyde. Glyceraldehyde activity was measured between 2-10mM to ensure that the maximum velocity of the reaction was recorded and that there was no apparent substrate inhibition.

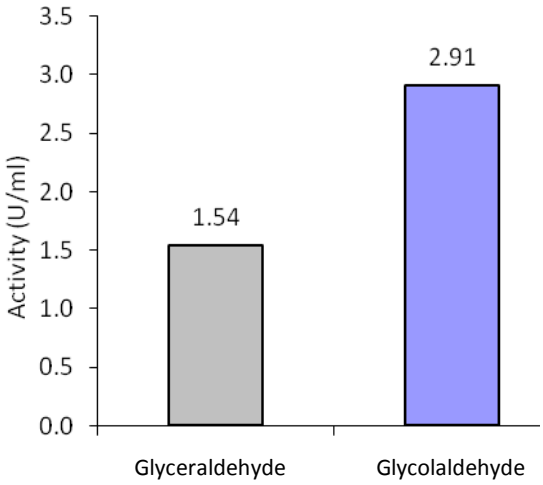


Fig. 3.16. Activity of aldehyde oxidoreductase with glyceraldehyde (5mM) and glycolaldehyde (10mM) as substrates One unit (U) of enzyme activity is defined as the amount of enzyme required to reduce 1µmol of dichlorophenolindophenol per minute.

Attempts were made to purify the enzyme further so that it could be conclusively identified. However, the protein rapidly lost activity when purified under aerobic conditions, concurring with data reported for the homologous enzyme in *S. acidocaldarius* (Kardinahl *et al*, 1999). It was therefore, not possible to purify the enzyme to homogeneity given the available resources.

3.3.7. Glycolate dehydrogenase (Glyoxylate reductase)

The conversion of glycolate to glyoxylate in this pathway is likely to be through the action of a glycolate dehydrogenase. Although the equilibrium lies towards the formation of glycolate, this step in the context of a catabolic pathway is still feasible as the subsequent enzymes in the citric acid cycle would utilise the glyoxylate formed (Chapter 1), driving the reaction forwards and allowing the pathway to proceed. *In vitro*, the enzyme assay is carried out in isolation with the subsequent enzymes absent; therefore the reaction had to be detected in the reverse direction.

Glyoxylate reductase activity was detected in cell extracts of *S. solfataricus* grown on C5 sugars as described in Materials and Methods. Unlike some of the previous enzymes investigated, there was no apparent inhibition at high substrate concentrations (Fig. 3.17). Kinetic parameters were determined using the Direct Linear method (Eisenthal & Cornish-Bowden, 1974): K_M = 5.0 (±0.2)mM, V_{max} = 1.5(±0.03) U/ml.

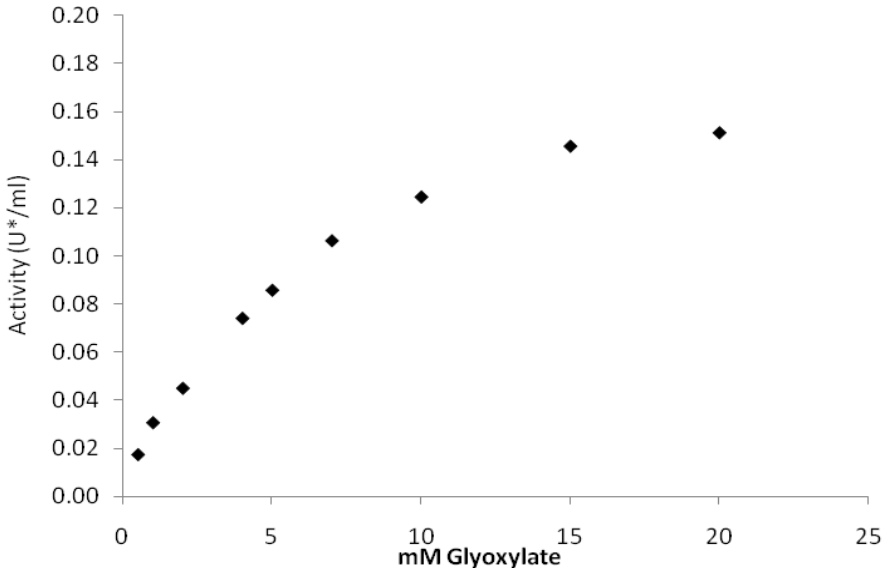


Fig. 3.17. Activity vs [S] plot of glyoxylate reductase activity in *S. solfataricus* cell extract from C5 grown cells assayed with 0-20mM glyoxylate and 0.2mM NADH. One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

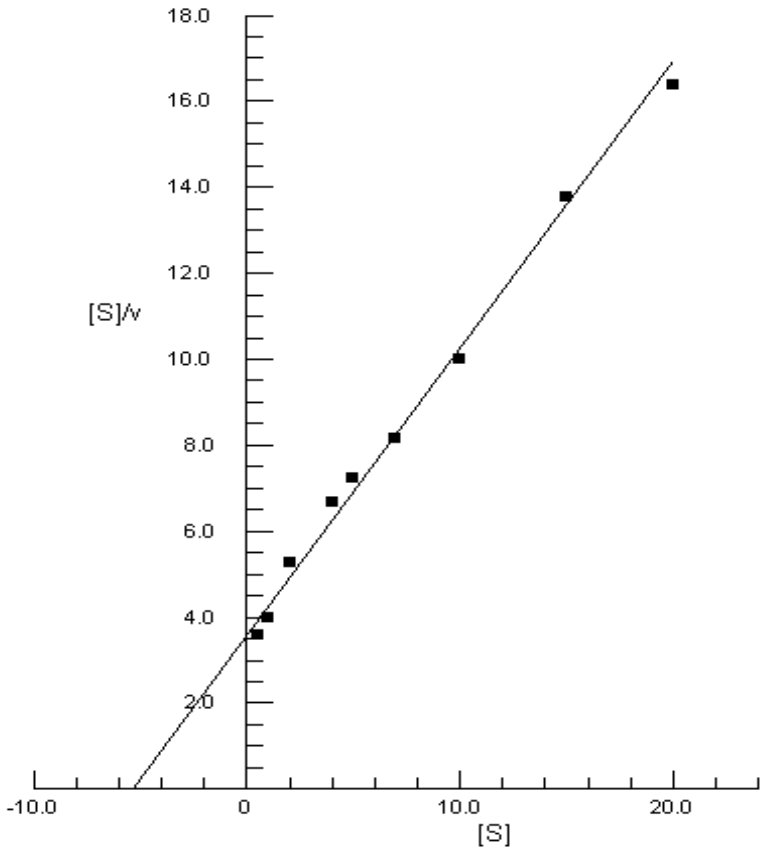


Fig. 18. Hanes - Woolf plot of glyoxylate reductase activity assayed in *S. solfataricus* cell extract from C5 grown cells with 0 – 20mM glyoxylate and 0.2mM NADH. [S] Is shown in mM and v in U/ml where one unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

Glyoxylate reductase appears to be constitutively expressed in *S. solfataricus* when grown on carbohydrates as a sole carbon source but activity was found to be increased in C5 grown cells (Fig. 3.01). Using the protein sequence of a previously characterised glyoxylate reductase enzyme from *Pyrococcus horikoshii* (Yoshikawa *et al*, 2006), a BLAST search was performed on the *S. solfataricus* genome. This revealed a candidate gene (SSO3187) that, when translated, shared 30% sequence identity at the amino acid level with the *P. horikoshii* enzyme and so is likely to account for the activity observed in the *Sulfolobus* cell extract.

3.3.8. Malate synthase and isocitrate lyase

Malate synthase and isocitrate lyase are both constitutively expressed in *S. acidocaldarius* (Uhrigshardt *et al*, 2002) and *S. solfataricus* grown on C6 sugars as a carbon source (Fig. 3.01). Interestingly, malate synthase activity is increased in C5 grown cells but no isocitrate lyase activity could be detected, indicating that expression of isocitrate lyase is suppressed when *S. solfataricus* is grown on C5 sugars.

Malate synthase was assayed in a C5 grown cell extract of *S. solfataricus* as described in Materials and Methods. There was no decrease in enzyme velocity at higher substrate concentrations and so kinetic parameters could be determined using the Direct Linear method (Eisenthal & Cornish-Bowden, 1974) (Table 4.03), and were found to be very similar to those reported for the *S. acidocaldarius* enzyme (Uhrigshardt et al, 2002).

Substrate	**K_M (µM)**	**V_{max} (U/ml)**
Glyoxylate	51.4 (±0.10)	1.9 (±0.004)
Acetyl-CoA	2.6 (±0.08)	2.4 (±0.008)

Table 3.06. Kinetic parameters of malate synthase with glyoxylate and acetyl-CoA at 70°C determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

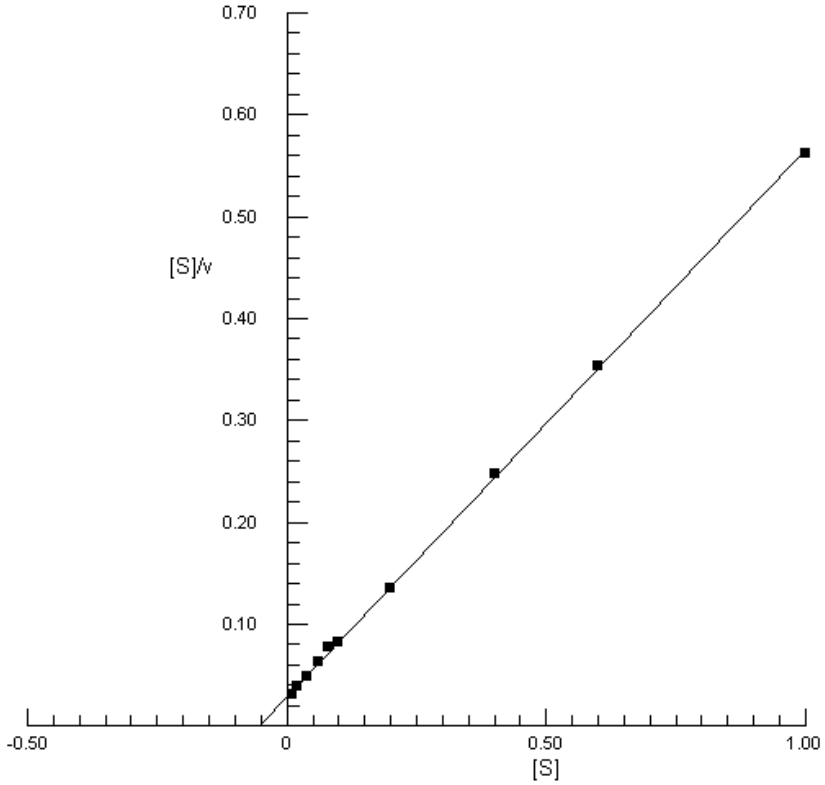


Fig. 3.19. Hanes-Woolf plot of malate synthase activity in *S. solfataricus* cell extract at 70°C with 0-1mM glyoxylate and 0.14mM Acetyl-CoA. [S] Is shown in mM and v in U/ml where one unit of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

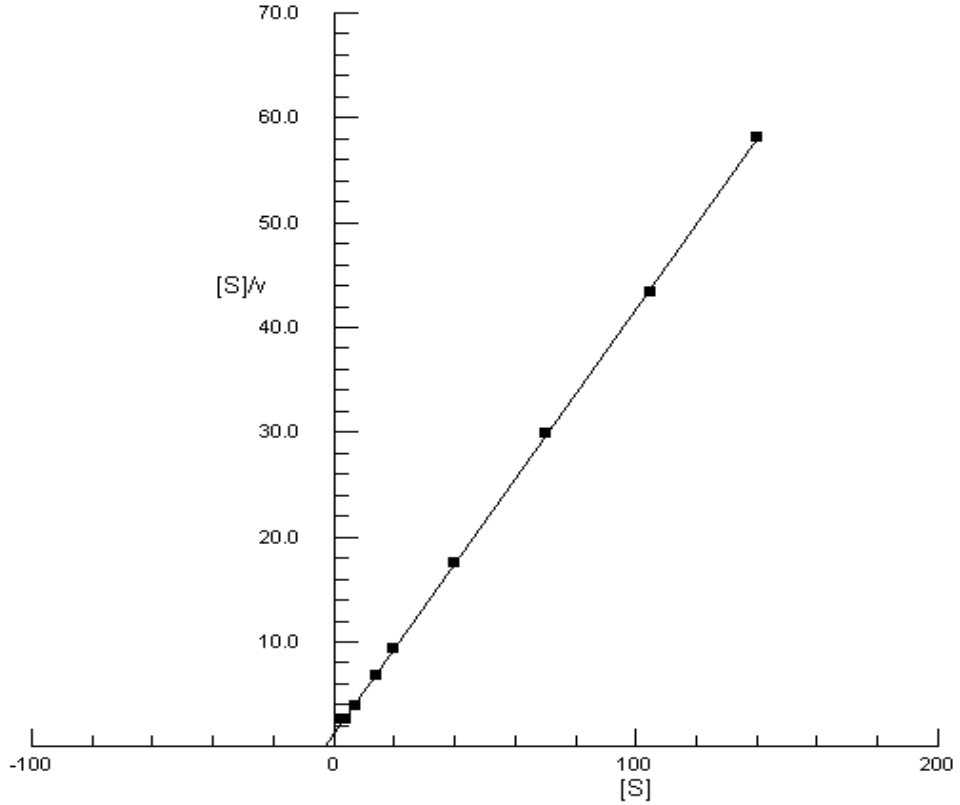


Fig. 3.20. Hanes-Woolf plot of malate synthase activity in *S. solfataricus* cell extract at 70°C with 0-140µM Acetyl-CoA and 10mM glyoxylate. [S] Is shown in mM and v in U/ml where one unit of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

Isocitrate lyase was assayed in a C6 grown cell extract of *S. solfataricus* as described in Materials and Methods. There was no observed decrease in enzyme velocity at high substrate concentrations allowing kinetic parameters to be determined using the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). For isocitrate V_{max} = 0.34 (±0.01) U/ml and K_M = 1.70 (±0.06)mM, 2 fold higher than the value reported for the *S. acidocaldarius* enzyme (Uhrigshardt *et al*, 2002).

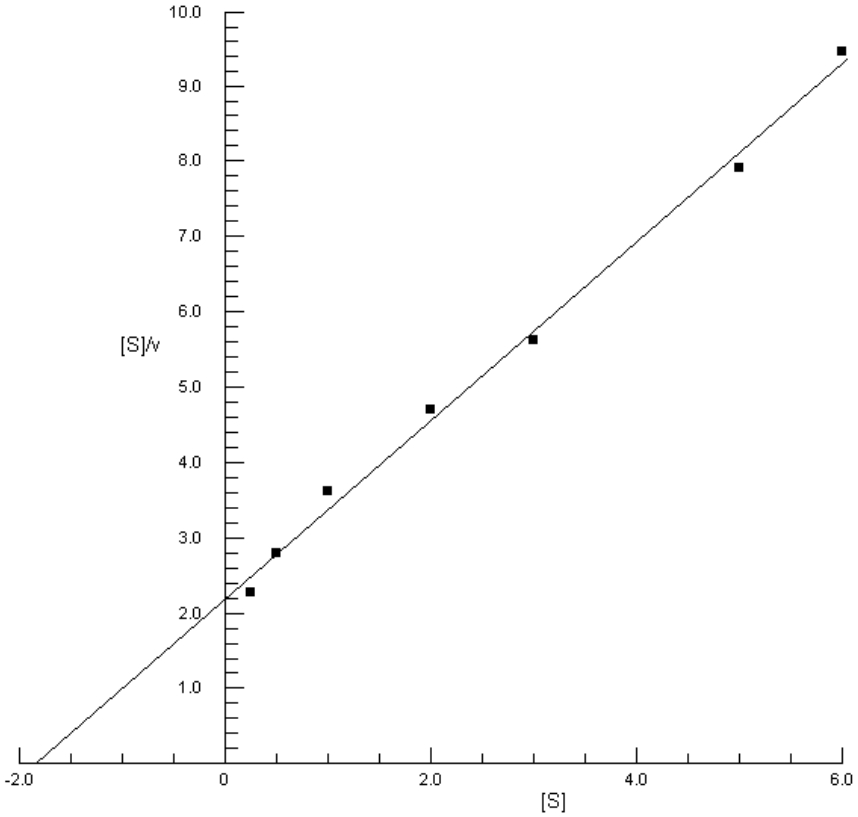


Fig. 3.21. Hanes-Woolf plot of isocitrate lyase activity in a *S. solfataricus* cell extract at 70°C with 0-6mM isocitrate and 2.5mM phenylhydrazine. [S] Is shown in mM and v in U/ml where one unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of glyoxylatephenylhydrazone per minute.

Putative genes for both of these enzymes are annotated in the genome of *S .solfataricus*, but unlike in bacteria (Yamamoto & Ishihama, 2003) and halophilic archaea (Serrano *et al*, 1998), they do not appear in an operon.

3.3.9. C5 metabolism genes in *S. solfataricus*

Using genome annotations and homology searches for each enzyme described above, candidate genes involved in the novel pathway proposed here were identified in the genome of *S. solfataricus*. A map was constructed (Fig. 3.22) showing the relative position and orientation of each gene involved in this pathway for C5 metabolism.

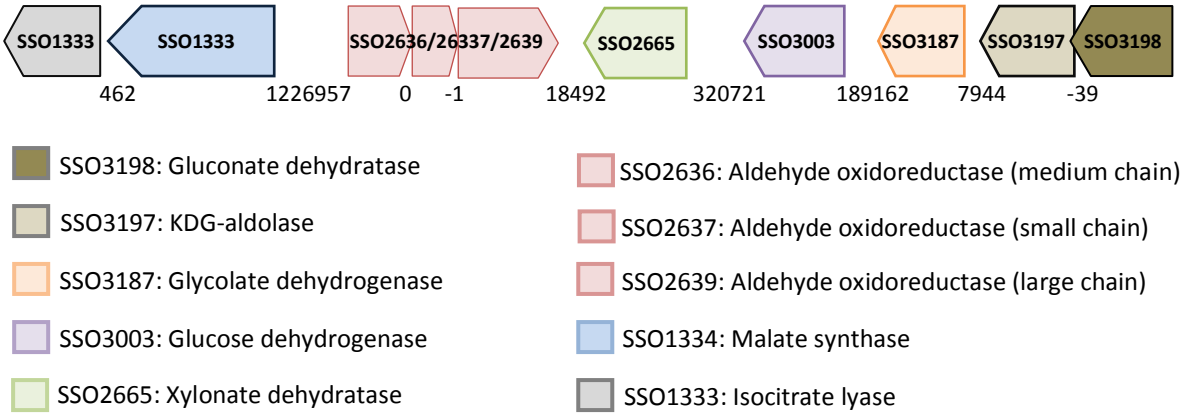


Fig. 3.22. Relative positions and orientations of genes involved in the new pathway for C5 metabolism in *S. solfataricus.* The distance between genes is shown in base pairs underneath the genes.

KDG-aldolase and gluconate dehydratase overlap by 39 bases with a putative Shine-Dalgarno sequence upstream of gluconate dehydratase suggesting these genes exist in an operon (Tolstrup *et al*, 2000). Both are constitutively expressed and show no detectable up regulation when *Sulfolobus* is grown on C6 or C5 sugars as a carbon source, indicating an alternative regulation mechanism. The aldehyde oxidoreductase subunits are predictably found together. The genes for the small and large chains of the enzyme have characteristic Shine-Dalgarno sequences upstream (appendix 1) indicating they exist in an operon (Tolstrup *et al*, 2000). The remainder of the genes involved in this pathway do not appear to be clustered, nor are they all up-regulated by growth on C5 sugars which suggests that their regulation is likely to be independent.

3.4. Discussion

A new scheme for C5 metabolism has been proposed (Chapter 1) predicting that *Sulfolobus* possesses a C5-specific dehydratase for the conversion of D-xylonate and L-arabinonate to 2-keto-3-deoxy-D-pentulosonic acid or 2-keto-3-deoxy-L-pentulosonic acid, which would then be cleaved into pyruvate and glycolaldehyde by the KDG-aldolase that operates in the C6 equivalent of this pathway. The enzymes glycolaldehyde oxidoreductase, glycolate dehydrogenase (glyoxylate reductase) and malate synthase would act in series to produce malate, which would then enter the citric acid cycle for complete oxidation to CO_2, NADH and $FADH_2$.

Activities of xylose dehydrogenase and KDG-aldolase did not differ between glucose and xylose-grown cells. This is not unexpected, as glucose dehydrogenase has equally high activity with D-glucose, D-galactose, D-xylose and L-arabinose. Furthermore, activities of glucose dehydrogenase and xylose dehydrogenase were found not to be additive in cell extracts of *S. solfataricus* (Lamble HJ, Thesis, 2005), indicating both sugars are oxidized by the same dehydrogenase.

The KDG-aldolase is also active with the 2-keto-3-deoxy-derivatives of these four sugars, so similar levels of activity would be expected under C5 and C6 growth. Activities of glycolate dehydrogenase and malate synthase were higher in cells grown on xylose than on glucose. This is expected as these enzymes are proposed to be involved in C5 sugar catabolism, but not C6. The C5 dehydratase showed similar levels of activity in glucose and xylose-grown cells, despite the enzyme having no detectable activity with gluconate or galactonate. This anomaly was also found with the previously identified gluconate dehydratase (Lamble HJ, Thesis, 2005), which again appears to be constitutively expressed in C5 grown cells even though it only catalyses the dehydration of C6 substrates.

The expression pattern of both dehydratases implies that an alternative mechanism of control must be responsible for the regulating the activity of these enzymes. Previous studies on the gluconate dehydratase (Kim & Lee, 2005), coupled with the data presented here suggest that these enzymes require phosphorylation for activity. Metabolic proteins requiring phosphorylation for activity have been reported in the model archaeon *Halobacterium salinarum* (Aivaliotis *et al*, 2009) and an active protein kinase from *S. solfataricus* has also been characterised (Lower & Kennelly, 2003), supporting

phosphorylation as a mode of regulation. The observed loss of activity at 37°C over time in the semi purified xylonate dehydratase is very unusual for a hyperthermophilic enzyme and may be enzymatically induced as a result of native *S. solfataricus* phosphatases co-purifying with the dehydratase. While the addition of a phosphatase inhibitor cocktail slowed this loss, it was still not sufficient to prolong the activity to allow sufficient purification of the enzyme for conclusive identification. The observed decrease in velocity at high substrate concentrations of both D-xylonate and L-arabinonate did not fit a classical substrate inhibition profile and was not a due to a change in pH as the reaction progressed. This type of inhibition was also reported for the D-arabinonate dehydratase characterised by Brouns *et al* (2006) at high substrate concentrations, but a reason for the decrease in velocity has not yet been found.

After investigating a number of possibilities, glycolaldehyde oxidoreductase was found to have optimum activity and in Bis-Tris buffer pH7.0 but no detectable activity in phosphate buffer at pH6.0 or 7.0 suggesting that the phosphate may have had some kind of inhibitory effect on the enzyme. The enzyme showed a decrease in velocity with increasing substrate concentration, but the results of fitting these data to a classical substrate inhibition model were inconclusive. This is possibly due to a combination of effects from substrate inhibition and a six-fold increase in observed background rate over the range of substrate concentrations shown in Fig. 3.15. This increased background rate will have resulted in less accurate data collection under these conditions, as the difference between the enzymatic and background rate reduces at higher substrate concentrations. While this enzyme could not be purified to homogeneity for identification, the estimated M_r from gel filtration combined with the presence of genes for all three subunits provides good evidence that the activity detected is that of the homologous enzyme to the glycolaldehyde oxidoreductase identified in *S. acidocaldarius* by Kardinahl *et al* (1999).

Glycolate dehydrogenase (glyoxylate reductase) activity was determined in the reverse direction owing to the position of the glycolate / glyoxylate equilibrium in solution. As previously discussed, this step in the pathway is still feasible *in vivo* as the subsequent enzymes in the glyoxylate and citric acid cycles would utilise the glyoxylate formed (chapter 1), driving the reaction forwards and allowing the pathway to proceed. A similar situation exists with an enzyme from the citric acid cycle, malate dehydrogenase. The equilibrium of this reaction favours the formation of malate, with an equilibrium constant of 2.33×10^{-12} M

and yet removal of oxaloacetate by the subsequent enzyme in the cycle, citrate synthase, causes this reaction to operate in the direction of oxaloacetate formation.

Malate synthase and isocitrate lyase are usually expressed together in aerobic micro-organisms growing on acetate as sole carbon source, the glyoxylate cycle being required for the biosynthesis of C4 compounds from the C2 nutrient. Significantly, in *Sulfolobus* growing on xylose and arabinose, malate synthase is produced in the absence of an isocitrate lyase, consistent with the metabolic scheme proposed (Fig. 3.02), although the latter enzyme was induced in cells grown on glucose. Uhrigshardt *et al* (2002), reported significant levels of acetate in *Sulfolobus acidocaldarius* growing on glucose, which may explain the induction of both glyoxylate cycle enzymes in the presence of this C6 substrate.

Malate synthase and isocitrate lyase are found 461 base pairs apart and TATA boxes as described by Reiter et al (1990) were found upstream of both genes. The large distance between the genes, coupled with the absence of a Shine-Dalgarno sequence upstream of isocitrate lyase, indicates that, unlike the situation in bacteria (Yamamoto & Ishihama, 2003) and halophilic archaea (Serrano *et al*, 1998), these genes do not exist in an operon in *S. solfataricus* (Tolstrup *et al*, 2000). The expression of one enzyme without the other under specific growth conditions has not been reported for any other organisms and so may be unique to *Sulfolobus* species utilising the pathway of C5 catabolism proposed here.

The positions of C5 metabolism genes in the *S. solfataricus* genome are not clustered in the way that is often found with inducible metabolic enzymes. This is unsurprising as all the enzymes involved are constitutively expressed, with the exception of isocitrate lyase as discussed previously. Interestingly, this constitutive expression of ED enzymes has also been reported in the thermoacidophilic euryarchaeon *Picrophilus torridus* (Reher *et al*, 2009). Being an opportunistic heterotroph having promiscuous metabolic enzymes constantly present offers an energetic advantage to *Sulfolobus* as substrates can be quickly metabolised without the need to induce expression of a specific set of enzymes depending on the carbon source available.

Overall these data support the novel metabolic pathway, proposed in Chapter 1, for the catabolism of the C5 sugars D-xylose and L-arabinose.

Chapter 4: Studies on recombinant enzymes from the modified C5 Entner-Doudoroff Pathway

4.1. Introduction

In the previous chapter, several enzymes were identified in the context of their role in C5 metabolism and the corresponding putative genes identified in the genome of *S. solfataricus*. It is important that the identities of these enzymes are confirmed, so that a better understanding of this new metabolic pathway can be achieved. Evidence that the correct gene products have been identified can be obtained by characterisation of recombinant enzymes.

Of these enzymes, xylonate dehydratase, glycolate dehydrogenase (glyoxylate reductase) and malate synthase were selected for further characterisation. The generation of a recombinant version of the molybdenum containing aldehyde oxidoreductase, (composed of three distinct subunits and inactivated in the presence of oxygen) was unfortunately not feasible given the time and resources available in this study. Further characterisation of these enzymes will not only increase the understanding of metabolic processes, but may also highlight new enzymes for potential industrial application.

4.1.1. Xylonate dehydratase

A gluconate dehydratase from *Sulfolobus solfataricus* has previously been purified from a native cell extract and the gene (SSO3198) identified through Mass Spectroscopy of protein derived fragments (Lamble *et al*, 2004). This enzyme was shown to catalyse the dehydration of gluconate to 2-keto-3-deoxygluconate in the non-phosphorylative Entner-Doudroff pathway. The substrate promiscuity for C6 and C5 substrates displayed by the glucose dehydrogenase and KDG aldolase from the same pathway led to the assumption that a single promiscuous enzyme catalysed the dehydration of gluconate to 2-keto-3-deoxy gluconate and xylonate to 2-keto-3-deoxylonate (2-keto-3-deoxy-D-pentulosonic acid). However, the native gluconate dehydratase enzyme was fully characterised and, although activity was shown with both D-gluconate and D-galactonate, no activity could be detected with the C5 equivalents D-xylonate or L-arabinonate. In the previous chapter, evidence was presented for a second, C5 specific dehydratase which is thought to be involved in a novel pathway for C5 metabolism in *S. solfataricus*.

A putative xylonate dehydratase gene (SS02665) was identified through a BLAST search of the *S. solfataricus* genome using the gluconate dehydratase gene sequence (SSO3198). The predicted protein showed a 62% sequence identity at amino acid level with the gluconate dehydratase and a predicted M_r of 44,700. Gluconate dehydratase was predicted to be a member of the enolase superfamily based on the presence of conserved residues involved in the general mechanism of α-proton abstraction (Lamble *et al*, 2004). An alignment of the predicted amino acid sequence of xylonate dehydratase with gluconate dehydratase was performed to identify those conserved catalytic residues in the xylonate dehydratase that would define it as a member of the enolase superfamily (Babbitt & Gerlt, 1997) (Fig. 4.01).

```
SSO2665 (XD)      MTKISEIEAYILG-KEVTSAQWASLMVLVRVTTNDGRVGWGETVSALRAEAVANFVKKIN
SSO3198 (GD)      -MRIREIEPIVLTSKEKGSATWASIMIVTRVITENGEVGYGEAVPTLRVISVYNAIKQVS
                    :* ***. :*  **  ** ***:*::.** *::*.**:**:*.:**. :* * :*::.

SSO2665 (XD)      TVLKGNDVFNVEKNRLEWYKHDFNMTISLESTTAYSAVDIASWDIIGKELGAPLYKLLGG
SSO3198 (GD)      KAYIGKEVEEVEKNYHEWYKQDFYLARSFESATAVSAIDIASWDIIGKELGAPIHKLLGG
                  ..  *::* :****  ****:** :: *:**:** **:****************::*****

SSO2665 (XD)      KTRDKVLVYANGWYQNCVKPEDFAEKAKEIVKMGYKALKFDPFGPYFNDISKKGLDIAEE
SSO3198 (GD)      KTRDRVPVYANGWYQDCVTPEEFAEKAKDVVKMGYKALKFDPFGPYYDWIDERGLREAEE
                  ****:* ********:**.**:******::*****************:: *.::**  ***

SSO2665 (XD)      RVKAVREAVGDNVDILIEHHGRFNANSAIMIAKRLEKYNPLFMEEPIHPEDVEGLRKYRN
SSO3198 (GD)      RVKAVREAVGDNVDILIKHHGRFNANSAIMIAKRLEKYNPGFMEEPVHHEDVIGLRKYKA
                  ************************************ *****:* *** *****:

SSO2665 (XD)      NTSLRIALGERIINKQQALYFMKEGLVDFLQADLYRIGGVTETKKVVGIAETFDVQMAFH
SSO3198 (GD)      STHLRVALGERLISEKETAFYVEEGLVNILQPDLTNIGGVTVGRSVIKIAEANDVEVAFH
                  .* **:*****:*.:::: ::::****::**.** .*****  :.*: ***: **::***

SSO2665 (XD)      NAQGPILNAVTLQFDAFIPNFLIQESFYDWFPSWKRELIYNGTPIDNGYAIIPERPGLGV
SSO3198 (GD)      NAFGSIQNAVEIQLSAVTQNLYLLENFYDWFPQWKRDLVYNETPVEGGHVKVPYKPGLGV
                  ** *.* *** :*:.*.  *: : *.******.***:*:** **::.*:. :* :*****

SSO2665 (XD)      EVNEKMLDSLKVKGEEYFNPEEPVWVVKGTWRDY--
SSO3198 (GD)      SINEKIIEQLRAEPIPLDVIEEPVWVVKGTWKNYGV
                  .:***:::.*:.:       ***********::*
```

Fig. 4.01. Sequence alignment of gluconate dehydratase (encoded by SSO3198) and putative xylonate dehydratase (encoded by SSO2665) from *S. solfataricus* showing conserved residues essential for catalysis; residues involved in Mg^{2+} co-ordination are highlighted in red, the electrophilic and acid/base catalysts are in grey, the His-Asp dyad is in green, and the electrophilic carboxylate group is in yellow.

Members of the enolase superfamily are structurally similar enzymes that catalyse different overall reactions but all perform the base catalysis of the α-proton removal from a carboxylate anion substrate. This reaction produces an enolate anion intermediate that is stabilised by the coordination of an essential Mg^{2+} ion (Rakus *et al*, 2008). The actual mechanism for this α-proton abstraction varies between the individual subgroups of the

superfamily depending on the functional groups that surround the intermediate in the active site (Babbitt & Gerlt, 1997).

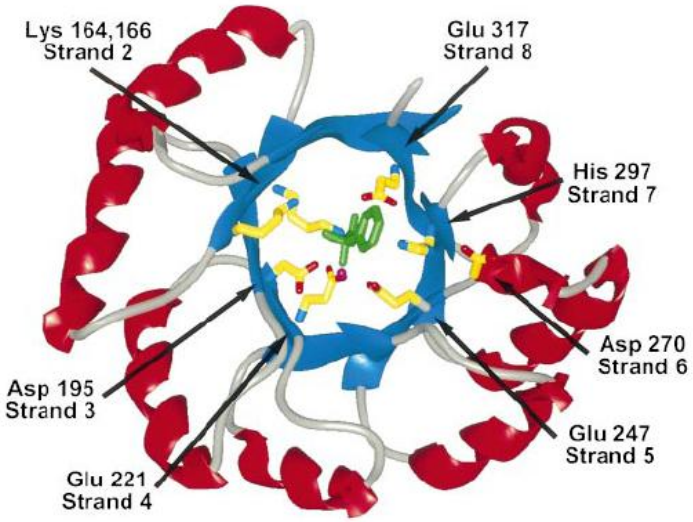


Fig. 4.02. The α/β barrel (TIM barrel) domain of mandelate racemase, a member of the enolase superfamily (based on Protein Data Bank entry 1mnr). Asp-195, Glu-221, and Glu-247 are metal ion ligands, Lys-166 and the Asp-270-His-297 dyad are general acid/base catalysts, and Lys-164 and Glu-317 are electrophilic catalysts. The β-strands surrounding the active site are numbered. Taken from Babbitt & Gerlt (1997).

The active site structure in all these enzymes (Fig. 4.02) is that of an α/β barrel in which each functional group involved in catalysis is projected by a different structural unit within the barrel domain. This distinction in functional groups around the same structural scaffold allows for the evolution of new catalytic activities while retaining those residues that participate in the partial reaction common to all the subgroups (Babbitt & Gerlt, 1997).

4.1.2. Glycolate dehydrogenase (glyoxylate reductase)

Glyoxylate reductase activity has been determined in a native cell extract of *S. solfataricus* and a putative gene identified in the *S. solfataricus* genome (SSO3187). The translated sequence shares 30% identity with the previously characterised glyoxylate reductase enzymes from *Thermococcus litoralis* (Ohshima *et al,* 2001) and *Pyrococcus horikoshii* (Yoshikawa *et al,* 2007). In *S. solfataricus*, the enzyme is thought to be involved in a novel metabolic pathway for C5 metabolism where it catalyses the oxidation of glycolate to glyoxylate which is then utilised via the glyoxylate and citric acid cycle. This enzyme has been implicated to have a role in metabolism in other hyperthermophilic archaea (Ohshima *et al,* 2001), although this has not yet been fully investigated.

Glyoxylate reductase belongs to the D-2-hydroxyacid dehydrogenase family (Ogino *et al*, 2008), which includes D-glycerate dehydrogenase (Goldberg *et al*, 1994), lactate dehydrogenase (Razeto *et al*, 2002), hydroxyisocaproate dehydrogenase (Dengler *et al*, 1997) and phosphoglycerate dehydrogenase (Schuller *et al*, 1995). The structure has been solved for glyoxylate reductase from *P. horikoshii* (Yoshikawa *et al*, 2007) where it is reported to function as a homodimer utilising NADH or NADPH as a cofactor. The residues for cofactor binding have been identified (a classic GXGXXG motif between residues 157-162 for $NAD(P)^+$ recognition and R181, which controls NADPH recognition) and are also present in the *S. solfataricus* enzyme, suggesting that it will also show dual specificity towards these cofactors (Fig. 4.03)

```
P. horikoshii      MKPKVFITREIPEVGIKMLEDEFEVEVWGDEKEIPREILLKKVKEVDALVTMLSERIDKE
T. litoralis       MKPKVFITRQIPENGIKMIEKFYEIELWKDPKAPPRGVLLEKVREVDALVTLVTDKVDKE
S. solfataricus    --MKIISTEKVPDECKNIIS------------VKDENITEEDYKNAEILLTWP-GRVNND
                      *:: *.::*:   :::.              . :  :. ::.: *:*    :::::

P. horikoshii      VFENAPKLRIVANYAVGYDNIDIEEATKRGIYVTNTPDVLTDATADLAFALLLATARHVV
T. litoralis       LLENAPKLKIIAQYAVGYDNIDIEEATKRGIYVTNTPGVLTDATADLAFALLLAVARRIV
S. solfataricus    LIGKMPNLKVIQTFSAGVDDLDFSIIPSH-VKVFSNAGAYSLSVAEHAWALILASAKGVG
                   :: : *:*:::  ::.* *::*:.  ..: : * ..... : :.*: *:**:** *: :

P. horikoshii      KGDRFVRSGEWKKRGVAWHPKWFLGYDVYGKTIGIIGLGRIGQAIAKR-AKGFNMRILYY
T. litoralis       EADAFVRSGEWKKSEVGWHPLMFLGYGLKGKTLGIVGFGRIGQALAKR-AKGFGMKIIYY
S. solfataricus    ------------------TKKRTIVYDVSEKTLLILGGGGIGSEVARIGKTAFRNYVIGI
                                     : *.:  **: *:* * **. :*:   ..*   ::

P. horikoshii      SRTRKEEVERELNAEFKPLEDLLRESDFVVLAVPLTRETYHLINEERLKLMKKTAILINI
T. litoralis       SRTRKPEAEEEIGAEYVDFETLLKESDFISLHVPLTKETYHMIGEKELKLMKPNAILINT
S. solfataricus    SRSFKKPEWFDERHGMTMLREKIGEADIIVDTLPLNKQTRGLLSYDLLKDIRRNAIIVNV
                   **: *     :       :.  : *:*::   :**.::*  ::. . ** :: .**::*

P. horikoshii      ARGKVVDTNALVKALKEGWIAGAGLDVFE----EEPYYNEELFKLDNVVLTPHIGSASFG
T. litoralis       SRGAVVDTNALIKALKEGWIAGAGLDVFE----EEPYYNEELFKLKNVVLAPHIGSATHE
S. solfataricus    GRGETVDEEGIYKLLKERQDVRFGTDVFWRKNGKEDFYNTKLWELDNFTGTLHT-AGAYG
                   .** .** :.: * ***   .  * ***     :* :** :*::*.*.. : *  :.:.

P. horikoshii      AREGMAELVAKNLIAFKRGEIPPTLVNREVIKIRKPGFE
T. litoralis       AREGMAELVAKNLIAFAKGEIPPNLVNKDVLTSSPP---
S. solfataricus    NESIMKRAMLIACLNVKK-YIDRGIADNEVRREDYV---
                    .. * . :    : . :  *   :.:.:*
```

Fig. 4.03. Alignment of archaeal glyoxylate reductase sequences. The classic GXGXXG motif for $NAD(P)^+$ recognition is present in all sequences. In *P. horikoshii,* arginine 181 is crucial for NADPH recognition (Yoshikawa, 2007) and is conserved in the *T. litoralis* enzyme which also utilises both co-factors (Ohshima *et al,* 2001).

All biochemically characterised glyoxylate reductase enzymes have been reported to have activity with both glyoxylate and hydroxypyruvate (Fauvart *et al*, 2007), but to date only glyoxylate reductase enzymes from *Rhizobium etli* and *T. litoralis* have been shown to have activity with additional substrates such as phenylpyruvate, α-ketobutyrate or pyruvate (Fauvart *et al*, 2007; Ohshima *et al*, 2001).

4.1.3. Malate synthase

Malate synthase is one of two key enzymes involved in bypassing the CO_2 evolving steps in the citric acid cycle. This glyoxylate shunt was identified by Kornberg and Krebs (1957) and the presence of malate synthase and the second key enzyme, isocitrate lyase, has been reported in all three domains of life (Chung *et al*, 1988; Serrano *et al*, 1998; Uhrigshardt *et al*, 2002). These two enzymes have been reported to exist in an operon in *E. coli* (Chung *et al*, 1988) and *H. volcanii* (Serrano *et al*, 1998) where there is evidence that they are co-transcribed in response to growth on acetate. Interestingly, this does not appear to be the case in the hyperthermophilic archaea *S. solfataricus* and *S. acidocaldarius*, where both enzymes are constitutively expressed when grown on glucose as a sole carbon source (Chapter 3; Uhrigshardt *et al*, 2002). Furthermore, recent work has shown that isocitrate lyase is actually repressed in *S. solfataricus* grown on C5 sugars and that the malate synthase and isocitrate lyase genes do not exist in an operon in the genome of *Sulfolobus* (Chapter 3) as they do in bacteria (Yamamoto & Ishihama, 2003) and halophilic archaea (Serrano *et al*, 1998).

Two malate synthase isoforms have been identified; isoform A which is found in fungi and plants, and isoform G which is found only in bacteria and is larger by approximately 200 residues (Lohman *et al*, 2008). Unusually, each isoforms is expressed in *E. coli* in response to specific growth conditions. The structure of both enzymes has been determined and a mechanism for the condensation of glyoxylate and acetyl-CoA has been proposed, similar to that of citrate synthase (Howard *et al*, 2000).

Work on *S. acidocaldarius* (Uhrigshardt *et al*, 2002) has shown that unlike all other malate synthases identified so far, the enzyme from this species is not dependent on Mg^{2+} for activity and that it has a subunit size larger than that of the bacterial isoform G. In *S. solfataricus*, and *S. acidocaldarius* malate synthase is thought to be involved in a novel non-phosphorylative pathway for C5 metabolism, in which the enzyme is regulated independently of isocitrate lyase. Further characterisation in terms of activity and structure of a recombinant malate synthase is important to investigate in this potentially new class of enzymes.

4.2. Materials and Methods

All chemicals were supplied by Sigma-Aldrich (Gillingham, UK) unless otherwise stated. Yeast extract, tryptone, agarose and agar were supplied by Melford (Suffolk, UK).

4.2.1. Growth Media

Luria Broth (LB)

10g tryptone, 5g yeast extract and 5g NaCl were dissolved in 1L of distilled water and autoclaved to sterilise before use.

YENB

7.5g yeast extract and 8g nutrient broth were dissolved in 1L of MilliQ water and autoclaved to sterilise before use.

SOC Medium

2.0g tryptone, 0.5g yeast extract, 10mM NaCl and 25mM KCl (final concentrations) were dissolved in 97ml distilled water, autoclaved and allowed to cool before the addition of 1ml of 2M Mg^{2+} stock (20.33g $MgCl_2.6H_20$, 29.65g $MgSO_4.7H_20$ in 100ml distilled water) and 1ml of 2M glucose. Both solutions were filter sterilised before use.

LB plates

15g agar was added to 1L LB medium and autoclaved to sterilise before use. This was allowed to cool to ~50°C before adding appropriate antibiotics and poured into plastic Petri dishes (Sterilin, Caerphilly, UK) to set. Plates were stored at +4°C for up to 2 weeks prior to use.

4.2.2. Gene cloning and expression of recombinant enzymes n *E. coli*

All genes were PCR-amplified from *S. solfataricus* genomic DNA, which was prepared using a standard DNA extraction kit (Qiagen, Crawley, UK), following manufacturer's instructions. PCR-amplification was carried out using Phusion™ (Finnzymes, Loughborough, UK) and an Eppendorf Mastercycler gradient PCR thermocycler (Eppendorf, Cambridge, UK) followed by A-tailing with *Taq* polymerase, unless otherwise stated. Amplified DNA fragments were purified by electrophoresis in a 1% (w/v) agarose gel, extracted using the Wizard Gel Extraction Kit (Promega, Southampton, UK) and ligated into the pGEMT-Easy vector (Novagen, Nottingham, UK) using T4 DNA ligase and 2x ligase buffer (Promega), unless

otherwise stated. After sequence determination to check PCR fidelity using the T7 forward and SP6 reverse primer sites, vectors were digested with the appropriate restriction enzymes, the inserts gel purified a second time and then ligated into pET vectors (Novagen) as detailed below. With the exception of D-xylonate dehydratase, restriction sites were selected to remove the His-tag from the vectors used.

Agarose gel for visualisation of DNA

A 50x TAE buffer stock was made by dissolving 242g Tris in 800ml distilled water and adding 100ml of 0.5M EDTA solution and 57.1ml glacial acetic acid, and the volume made up to 1L with distilled water. 0.5g Agarose was added to 50ml 1x TAE Buffer and heated in a microwave until the agarose was completely dissolved. The solution was allowed to cool before adding 2.5µl ethidium bromide (0.5µl / 10ml). This 1% molten gel was poured into the gel cassette; the lane separator positioned, and allowed to set. The gel was then placed in a gel tank, covered with 1X TAE buffer and run at 80-90v after the addition of DNA samples mixed with 6x loading dye (NEB, Herts, UK).

A-Tailing blunt PCR products

1-7µl of PCR product, 1µl of *Taq* 10X buffer (with $MgCl_2$), 0.2mM dATP and sterile MilliQ water were added to a final volume of 10µl in a 0.2ml thin-walled PCR tube (Biohit, Devon, UK). 0.5µl *Taq* polymerase was then added and the reaction incubated at 70°C for 30 min in an Eppendorf Mastercycler gradient PCR machine. The reaction was then cooled on ice for 15 min and 1-3µl used in the ligation reaction.

Ligation of DNA into pGEMT / pET vectors

A reaction containing 1-3µl of A-tailed PCR product, 1µl pGEMT, 5µl 2x ligation buffer and 0.5µl DNA ligase (Promega) and made up to 10µl with sterile MilliQ water was incubated overnight at 4°C before ethanol precipitation and transformation into JM109 competent cells. pET vector ligations were set up in the same way using 1µl of the digested pET vector and 1µl of 10x ligation buffer (in place of the 2x buffer).

Ethanol precipitation of ligations

20µl Sterile water was added to 10µl ligation reaction and the remaining ligase enzyme heat killed by incubating at 70°C for 20 min. 3µl (0.1 volumes) of 2.5M sodium acetate (pH5.2) were added followed by 60µl (2 volumes) of 100% ethanol. This was inverted

several times to mix and incubated at -20°C for 30 min before centrifugation at 13683 x g for 20 min. The supernatant was removed and 100µl of 70% ethanol added to remove traces of salt. This was inverted to mix and centrifuged at 13683 x g for 10 minutes. The supernatant was removed and the pellet allowed to air dry before it was resuspended in sterile MilliQ water.

Preparation and transformation of electrocompetent cells

A single colony picked from a stock plate of JM109 (Stratagene, Leicester, UK) or Rosetta (Novagen) cells was grown in 10ml YENB medium overnight. 1ml of this culture was used to inoculate 250ml YENB and grown to an OD_{600} of ~1.0 at 37°C with shaking. Cells were cooled on ice before centrifugation at 5000 x g for 10 min. Each pellet was resuspended in 25ml ice cold MilliQ water and then pooled into a single fraction before again being centrifuged at 5000 x g for 10 min. The supernatant was discarded and the pellet resuspended in 25ml MilliQ water per original 250ml of culture. The suspension was centrifuged at 5000 x g for 10 min and the supernatant discarded. The pellet was resuspended in 5ml 10% (v/v) glycerol per original 250ml pellet. The pellet was resuspended in 750µl 10% (v/v) glycerol per original 250ml pellet and dispensed into pre chilled 0.5ml tubes in 40µl aliquots. Cells were transformed using Genepuler™ cuvettes of 10mm pathlength and a Micropulser™ electroporator (Biorad, HERTS, UK) using the Ec1 setting for *E. coli* (1.8kV with a time constant of approximately 5 ms).

4.2.3. Cloning of individual genes

D-xylonate dehydratase

A putative dehydratase gene (SS02665) was identified through a BLAST search of the *S. solfataricus* genome using the gluconate dehydratase gene sequence (SSO3198). The gene was PCR amplified using the primers catatgacaaagatttcagaaatcgaa (forward) and ggatccatagtctctccaagttcctttaac (reverse) and cloned into pET3a and pET19b vectors (Novagen, Nottingham, UK) using the *Nde*1 and *Bam*H1 sites to produce both untagged and His-tagged constructs. The plasmid was transformed into a series of host *E. coli* strains including BL21(DE3), KRX, Rosetta™, Arctic Express™, C41 and C43. Strains containing either of these constructs were grown in LB or Overnight Express™ media (Novagen) containing appropriate antibiotics. LB cultures were grown with and without IPTG induction and at either 37°C or 45°C following induction in Rosetta. All expression cultures

were supplemented with 20mM $MgCl_2$. A cell extract was prepared as described below but with the addition of 20mM $MgCl_2$ to the lysis buffer.

Glycolate dehydrogenase (glyoxylate reductase)

A putative glyoxylate reductase gene (SSO3187) was identified in the genome of *S. solfataricus* using a BLAST search with the glyoxylate reductase sequence from *Pyrococcus horikoshii* (Yoshikawa 2007). The gene was PCR amplified using the primers catatgagagtctgaaacttaacg (forward) and ggatcctcctctatttgtttatctacc (reverse) and cloned into pET3a using the *Nde*1 and *Bam*H1 sites, shown underlined. Expression of the recombinant enzyme was carried out in Rosetta™ grown at 37°C overnight in Overnight Express Medium™. A cell extract was prepared as described below.

Malate synthase

The putative malate synthase gene (SSO1334) annotated in the genome of *S. solfataricus* was PCR amplified and cloned directly into pET19b using the primers; ccatgggatcaaatgtcttctagcc (forward) and ctcgagcctatttaccagtaattgtcg (reverse); the *Nco*1 and *Xho*1 restriction sites used for cloning are underlined. The construct was sequenced using the T7forward and T7 Reverse primer sites present up and downstream of the insert, plus a custom internal primer (ctgtggatatttcgtgagagactc) to obtain a complete read. The gene was expressed in Rosetta™ grown at 37°C overnight in Overnight Express Medium to produce ~50% soluble recombinant enzyme. A cell extract was prepared as described below.

4.2.4. Preparation of cell extracts

Unless otherwise stated, cells were harvested by centrifugation at 5000 x g for 10 min and resuspended at 0.2g/ml of 50mM Tris-HCl, pH8.0. After 10 min at room temperature, cells were lysed using a cell disruptor (One shot model, Constant Systems, Warwick, UK). Cell debris was removed by centrifugation at 12520 x g for 30 min. This cell extract was then incubated at 70°C for 10 min to remove the majority of *E. coli* proteins unless otherwise stated. In the majority of cases, this heat treatment produced a cell extract which was 80-90% pure and therefore sufficient for accurate enzyme assays.

4.2.5. Further purifications

Heat precipitated cell extracts were subjected to further purification by FPLC (ÄKTA, GE Healthcare, Bucks, UK). Samples were filtered through a 0.22μm filter (Millipore, Herts, UK) before each purification step. Anion exchange chromatography was performed using two 5ml HiTrap Q Sepharose columns (GE Healthcare, Bucks, UK) using 50mM Tris / HCl (pH8.5) with an elution gradient of 0-1M NaCl and a flow rate of 1.5ml / min. Fractions containing the highest activity were then pooled and concentrated by centrifuging samples at 2880 x g for 5 min using an Amicon Ultra-40 centrifugal tube (Millipore). Gel filtration was carried out on a Superdex 200 10/300GL (10mm x 300mm) column (GE Healthcare) using 50mM Tris / HCl (pH8.0) with a flow rate of 0.5ml / min.

4.2.6. Glycolate dehydrogenase / malate synthase coupled enzyme assay

Assays were performed at 70°C in 1ml of 50mM sodium phosphate buffer, pH6.0. The assay mixture contained 0.5mM NAD^+, 10mM glycolate, 0.14mM acetyl-CoA, and 20μg pure recombinant malate synthase enzyme. The reaction was started by the addition of 10μg glyoxylate reductase. The reduction of NAD^+ was followed by the increase in absorbance at 340nm (ε = 6220 $M^{-1}cm^{-1}$). One unit of enzyme activity is defined as the amount of enzyme required to reduce 1μmol of NAD^+ per minute.

4.2.7. Homologous expression of xylonate dehydratase in *S. solfataricus* M16

Homologous expression of the putative xylonate dehydratase was achieved using the method of Albers *et al* (2006), and is summarised below (Fig. 4.04). The *Sulfolobus solfataricus* uracil auxotroph strain (ΔpyrEF/lacS double mutant) was kindly provided by Dr. Sonja-Verena Albers (Max Planck Institute for Terrestrial Microbiology, Germany).

The putative xylonate dehydratase gene was PCR amplified using the primers ccatggtgacaaagatttcagaaatcgaa (forward) and ggatccatagtctctccaagttcctttaac (reverse) and cloned into the vector pMZ1 using the restriction sites *Nco*1 and *Bam*H1. This vector contains an upstream D-arabinose inducible promoter and downstream His and Strep tags either side of the gene insertion site. The promoter, gene and tags were then excised from the vector as a single piece of DNA using the restriction sites *Avr*II and *Eag*1, and ligated into the modified viral vector pMJ0503. This vector contains a pUC18 region to allow propagation in *E. coli* before transformation into *Sulfolobus,* and a selectable pyrEF gene which complements the *Sulfolobus* uracil auxotroph mutant, allowing transformants to be

grown under selective conditions. After ligation, the pMJ0503 vector containing the insert was transformed into the Stbl4 strain of *E. coli* engineered for the propagation of large vectors (Invitrogen). Successfully transformed cells were picked and grown in LB supplemented with 0.5% glucose. Plasmids were extracted using a Miniprep Wizard kit (Promega) and transformed into *Sulfolobus*, where cells were grown on Gelright™ plates at 70°C for 5-7 days to allow successful transformants to grow. Single cells were picked and grown for 8 days on selective solid media at 70°C. After this growth period, a loop of cells were transferred to 10ml selective liquid medium and grown for 2 days. Three of these cultures were then used to inoculate three larger cultures of 500ml of Brock's medium (pH3.5) containing either no arabinose, 0.2% L-arabinose or 0.2% D-arabinose. All cultures were grown at 75°C for two days before harvesting the cells.

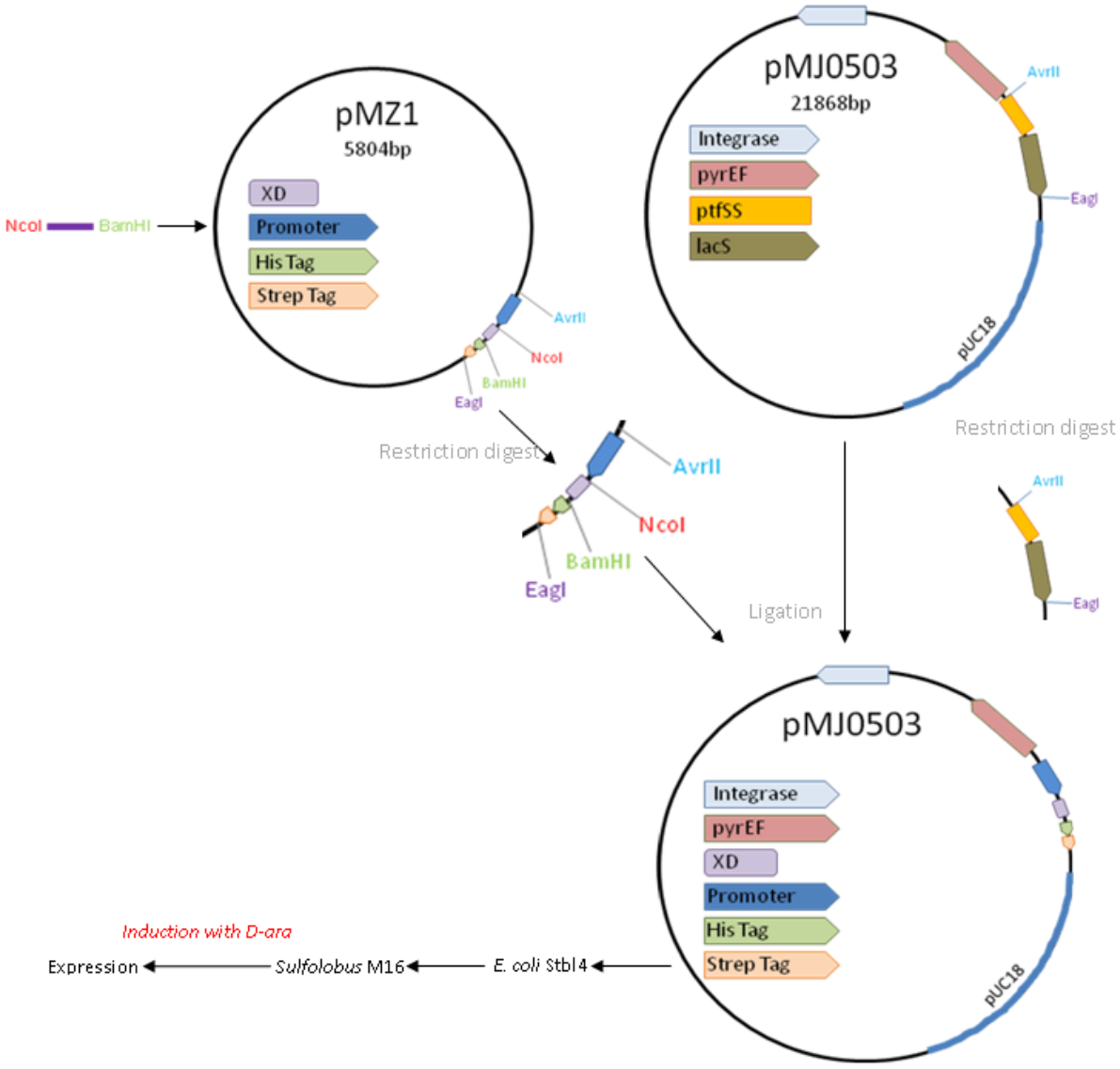


Fig. 4.04 Overview of *S. solfataricus* homologous expression system.

Vector	Forward primer	Reverse primer
pMZ1	caatatgttaacaaaacgtcttttac	gattaaaagaaaactcatttggtatag
pMJ0503	ctgataccattaggatgctaaac	gtatgttgtgtggaattgtg

Table 4.01. Vector specific sequencing primers for pMZ1 and pMJ0503 vectors used in the *S. solfataricus* homologous expression system.

***Sulfolobus solfataricus* M16 medium**

Brock's medium (pg. 15) was supplemented with 0.1% tryptone and +/- 10µg/ml uracil for the untransformed and transformed cells, respectively.

***Sulfolobus solfataricus* M16 plates**

500ml of double concentration Brock's medium, adjusted to pH 3.5, containing 6 ml of 0.5M $CaCl_2$ and 10 ml of 1M $MgCl_2$ was warmed to 60^{o}C. 6.4 g of Gelrite^{TM} was melted in 500 ml H_2O by microwaving. The two solutions were combined and poured into autoclaved glass Petri dishes. Plates were stored at 4°C until required.

Electroporation of Electromax Stbl4 cells (Invitrogen)

Cells were thawed on ice prior to use. 1µl of ligation mix was mixed with 20µl of electrocompetent cells, and incubated on ice for 1 min before transferring the cell/DNA mix to a cold cuvette. Cells were pulsed once with 25 µF, 2.5 kV and 200 Ω with the time constant between 4 to 5.2 ms. 1 ml of SOC medium was added immediately and the cells transferred to a 50ml Falcon tube and grown for 1.5 h at 30^{o}C with shaking. Cells were plated out and grown grow at 30^{o}C. Single colonies were picked and grown overnight, and plasmids extracted using a Miniprep Wizard kit (Promega).

Luria Broth (LB) and agar plates

Filter sterilised glucose solution was added to both LB growth medium and LB agar plates to give a final concentration of 0.5%.

Electroporation of *Sulfolobus solfataricus* M16

Two cultures of *Sulfolobus solfataricus* M16 (ΔpyrEF/lacS double mutant) were started to ensure that the negative control without uracil did not grow over an OD_{600} of 0.3. After 2 days growth, a 500ml culture was inoculated and allowed to grow overnight to an OD_{600} of between 0.2-0.3. The overnight culture was cooled on ice, and centrifuged at 4000 x g at 4°C for 20 min. The cell pellet was resuspended in 1ml cold sucrose water (6.9ml 20%

sucrose in 200 ml water) and then diluted into a total of 50ml sucrose water and centrifuged again. This was repeated with the pellet resuspended in a total of 10ml sucrose water and centrifuged again. The cell number was adjusted to 10^{10} cells/ml (e.g. If the OD_{600} of a 50ml culture is 0.1 the cells would be resuspended in 200µl sucrose or if original OD_{600} is 0.15 then add 300µl of sucrose). Cells were incubated on ice before electroporation. 0.5µl Vector was mixed with 50µl of cells, and transferred to a chilled cuvette. Cells were electroporated at 1.5kV, 25µF and 400Ω, with a time constant of approximately 10 ms. After electroporation 1ml of medium + uracil were added and the cells transferred to a 1.5ml Eppendorf. Cells were regenerated at 75°C for 1hr and the cells aerated every 20mins by opening the tube. 50ml pre-warmed medium plus uracil was inoculated with the transformed cells and grown for 2-3 days until the cells reached an OD_{600} of 0.5. After this time, cells were transferred to selective medium without uracil and grown for 2 days. After growth, cells were plated onto Gelrite™ plates and incubated for 6-8 days. Cells were picked and grown in Brock's medium containing 0.2% D-arabinose for expression.

4.2.8. Nickel affinity chromatography

2ml of metal chelating cellulose (Bioline, London, UK) was loaded into a Poly-Prep chromatography column (Biorad) and washed with 10ml MilliQ water. 5ml of 400mM $NiSO_4$ solution was then passed down the column, followed by a 10ml wash of 50mM Tris-HCl buffer (pH8.0) containing 5mM Imidazole (Acros Organics, Geel, Belgium) and 10mM $MgCl_2$. A cell extract was then passed down the column and the flow through collected in 1ml fractions. Bound proteins were eluted by sequentially washing the column with 1ml volumes of Tris-HCl buffer (pH8.0) containing 50-1000mM imidazole and 10mM $MgCl_2$.

4.3. Results

4.3.1. D-Xylonate dehydratase

4.3.1.1. Gene cloning and heterologous expression

The putative gene was PCR amplified (Fig. 4.05) producing a fragment of the size predicted from the *S. solfataricus* genome. It was cloned into both His-tagged (pET19b) and non-tagged vectors (pET3a), transformed into several different *E. coli* hosts and expressed under a variety of conditions as described.

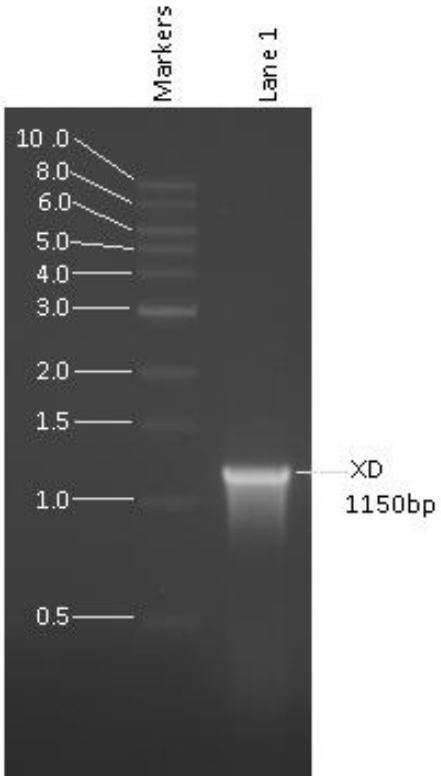


Fig. 4.05. 1% Agarose gel visualised with ethidium bromide under UV light. Lane 1 - PCR amplification of the putative xylonate dehydratase (XD) gene from *S. solfataricus*. The PCR product was the expected size of 1.2kb. A 1kb ladder (NEB, Massachusetts, USA) with sizes in kb is shown to the left of the figure.

Despite utilising a wide range of conditions, none of the expression trials yielded a soluble and active recombinant protein. Large amounts of insoluble protein were observed when the tagged construct was expressed in Rosetta™ grown in Overnight Express™ medium, but no over-expression of soluble or insoluble recombinant protein could be seen in the untagged construct under the same conditions (Fig. 4.06).

IPTG controlled expression in Rosetta™ with increased temperature cultivation, as described by Koma *et al* (2006), also failed to produce any detectable over-expressed protein in either the soluble or insoluble fraction. The constructs were also transformed into the *E. coli* strain Arctic Express, which is designed to increase solubility of recombinant proteins by inducing chaperonin proteins at low growth temperatures. *E. coli* C41 and C43, reported for their ability to express toxic proteins (Dumon-Seignovert *et al*, 2004), were also utilised in an attempt to produce soluble protein, although none of these conditions proved to be successful.

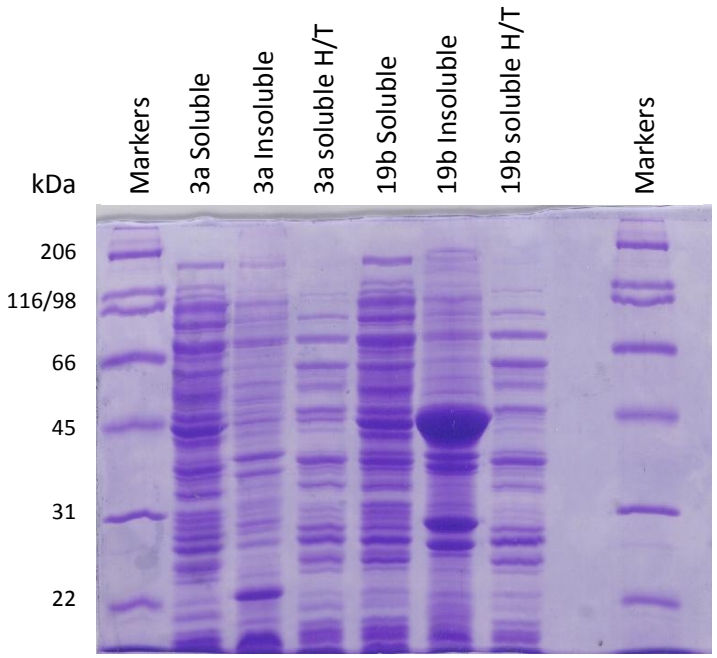


Fig. 4.06. SDS PAGE gel showing the expression trials of both untagged and tagged xylonate dehydratase constructs in Rosetta with Overnight Express™ medium. Samples were taken 24h after induction and prepared as described in Materials and Methods. Soluble fractions were heat treated (H/T) at 70°C for 10 min.

Despite no visual indication of expression by analysis of fractions on an SDS PAGE gel, all soluble fractions were assayed for activity as described and in the case of the tagged constructs, subjected to purification by nickel affinity chromatography as described in Materials and Methods. An example of results from the nickel affinity purification can be seen in Fig. 4.07. It is clear from the SDS PAGE gel that the elution fractions contain no bound proteins indicating that either no soluble protein had been expressed, or that the His-tag has become internalised and so was not available to interact with the column.

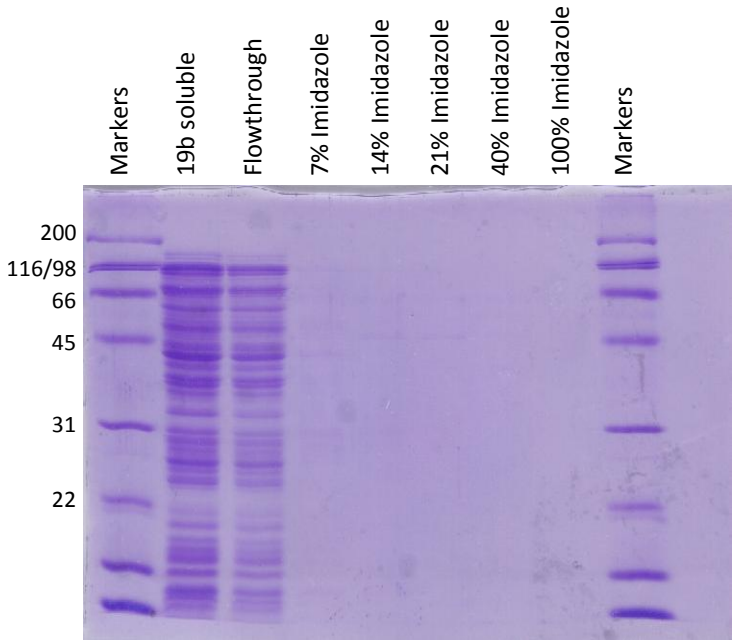


Fig. 4.07. SDS PAGE gel showing the soluble sample from a Rosetta™ expression trial passed through a nickel affinity column and eluted with increasing amounts of imidazole as described. The sizes of the Broad-Range markers (Biorad) are shown to the left of the figure in kDa.

4.3.1.2. Gene cloning and homologous expression

Due to the difficulties encountered in obtaining soluble recombinant protein by conventional heterologous expression methods, attention was redirected to a new homologous expression system developed by Albers *et al* (2006), as described in Materials and Methods.

New primers were designed to incorporate the desired *Nco*1 and *Bam*H1 restriction sites, and the putative xylonate dehydratase gene amplified with Phusion™ polymerase as described. The PCR product was ~1.2kb (Fig. 4.08), as predicted by the gene sequence and was cloned into pGEMT for sequencing, which confirmed the fidelity of the PCR reaction.

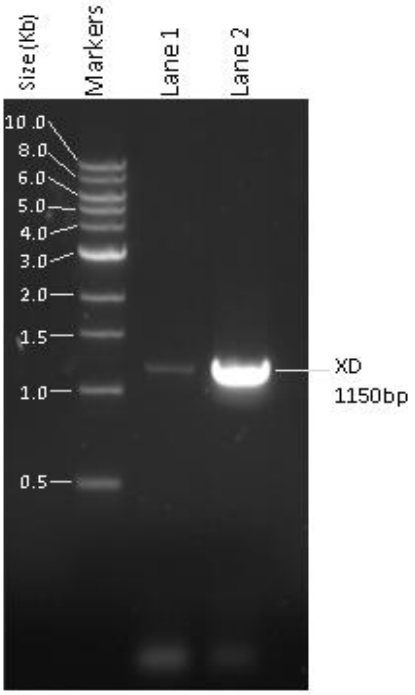


Fig. 4.08. 1% Agarose gel visualised with ethidium bromide under UV light. Lanes 1 and 2 - PCR amplification of the xylonate dehydratase (XD) gene from *S. solfataricus* at annealing temperatures of 56.0°C and 52.4°C respectively. The PCR product is ~1.2kb as predicted from the gene sequence. Marker sizes are shown in kb on the left of the figure.

The pGEMT construct and pMZ1 vector were digested with *Nco*1 and *Bam*H1 restriction enzymes, as shown in Fig. 4.09. The insert and pMZ1 vector were gel purified and ligated as described.

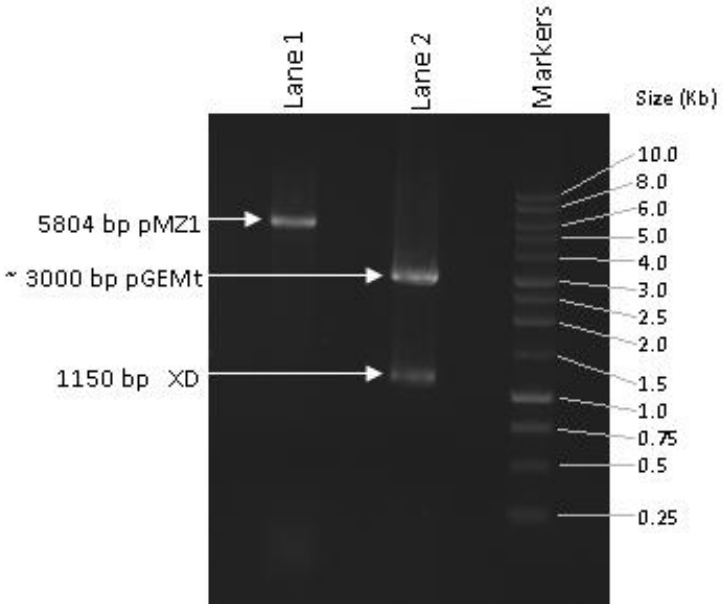


Fig. 4.09. 1% Agarose gel visualised with ethidium bromide under UV light. Lane 1 - pMZ1 vector linearised with *Nco*1 and *Bam*H1. Lane 2 – pGEMT containing xylonate dehydratase (XD) gene digested with the same restriction enzymes showing a band of the expected size. Marker sizes are shown in kb on the right of the figure.

After propagation in JM109 cells, plasmids were purified from successful transformants and restriction digests were performed using either *Nco*1 / *Bam*H1 (Fig. 4.10) or *Avr*II / *Eag*1 (Fig. 4.11) to confirm the presence of the correct insert. Sequencing with vector specific primers was also performed to confirm that the upstream promoter and downstream tags were present (appendix 2).

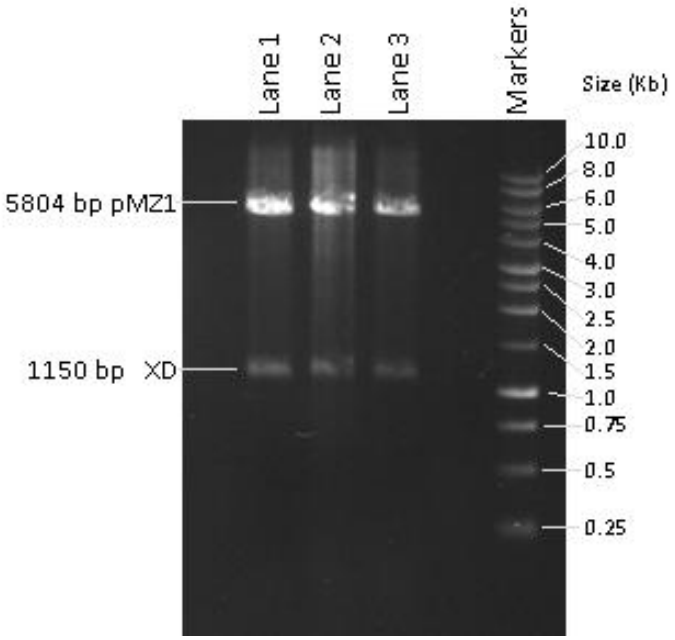


Fig. 4.10 – 1% Agarose gel visualised with ethidium bromide under UV light. Lanes 1 – 3 show three clones of the pMZ1 / xylonate dehydratase (XD) construct digested with *Nco*1 and *Bam*H1 to confirm the presence of the correct size insert. Marker sizes are shown in kb on the right of the figure.

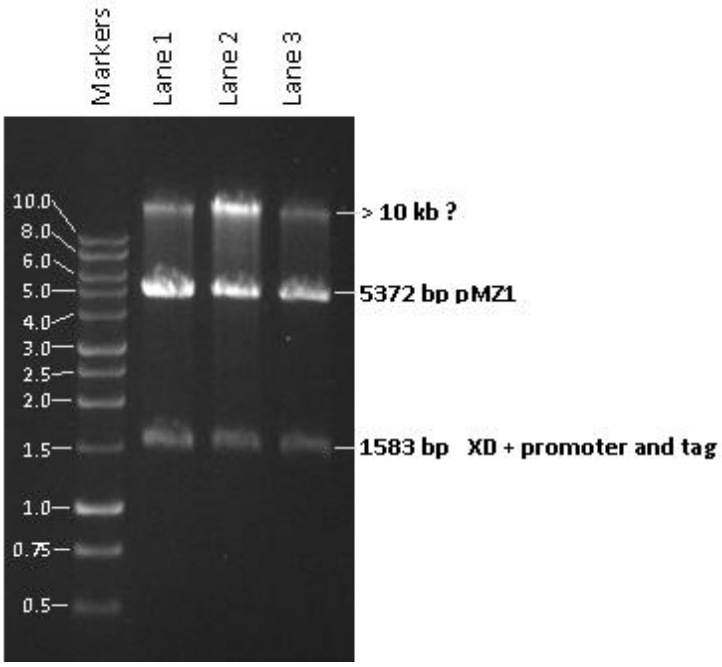


Fig. 4.11 – 1% agarose gel visualised with ethidium bromide under UV light. Lanes 1 – 3 show three clones of the pMZ1 / XD construct digested with *AvrII* and *Eag1* to confirm the presence of the correct size insert plus promoter and tags. Marker sizes are shown in kb on the right of the figure.

pMZ1 clones digested with *AvrII* and *Eag1* had a third band present on the agarose gel. This was a large band over 10kb in size and did not correspond to either a product of the digestion or any partially digested vector and insert. However, the insert was of the expected size for the xylonate dehydratase gene plus the flanking promoter upstream and affinity tags downstream and so was purified and ligated into pMJ05-03 which had also been digested with *Avr*II and *Eag*1. pMJ05-03 ligations were transformed into *E. coli* stbl4 electromax cells (Invitrogen) and the presence of the insert was confirmed by restriction digestion (Fig. 4.12) and sequencing (appendix 2) using vector specific primers.

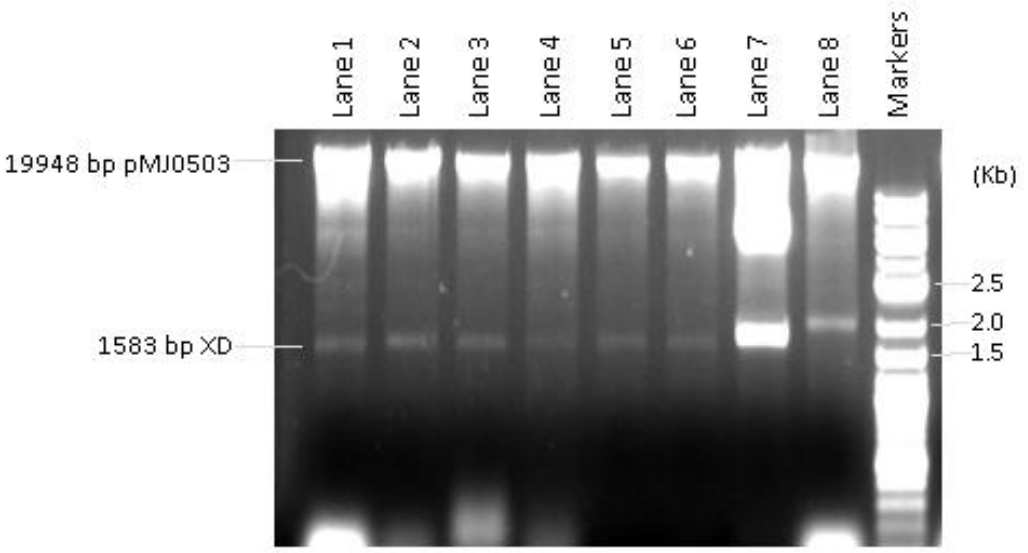


Fig. 4.12 – 1% Agarose gel visualised with ethidium bromide under UV light. Lanes 1 – 6 show multiple clones of the pMJ0503 / XD construct digested with *Avr*II and *Eag*1 to confirm the presence of the correct size insert plus promoter and tags. Lane 7 contains a positive control for the digestion (E3 gene in pMZ1), lane 8 shows the lacS gene cut out of pMJ0503 after restriction with *Avr*II and *Eag*1. Marker sizes are shown in kb on the right of the figure (due to the overexposure of the gel it is not possible to identify the larger and smaller fragment sizes).

Electrocompetent *S. solfataricus* cells were prepared and transformed as described and cell extracts made as described in Materials and Methods. Enzyme activity was determined using the standard assay described in general Materials and Methods with xylonate as a substrate. A low dehydratase activity could be detected in the uninduced and L-arabinose grown samples due to the presence of the constitutively expressed native enzyme. The specific activity was, however, almost twofold higher in the D-arabinose induced sample (Table 4.02).

Sample	A_{549} min^{-1}	U/ml	Protein concentration (mg/ml)	Specific activity (U/mg)
Uninduced	0.04	0.72	7.1	**0.10**
L-Ara induced	0.01	1.80	17.0	**0.11**
D-Ara induced	0.13	2.33	13.0	**0.18**

Table 4.02. Xylonate dehydratase activity in cell extracts from 3 expression cultures under standard assay conditions at 70°C. Protein concentrations were determined by the Bradford method (Bradford, 1976). One unit (U) of enzyme activity is defined as the amount of enzyme required to form of 1μmol of 2-keto-3-deoxypentulosonic acid per min.

The soluble cell extract form the D-arabinose induced sample was subjected to nickel affinity chromatography (Fig. 4.13) as described in Materials and Methods. Fractions containing protein species of the expected size (determined by SDS PAGE) were assayed under standard conditions but showed only a low dehydratase activity.

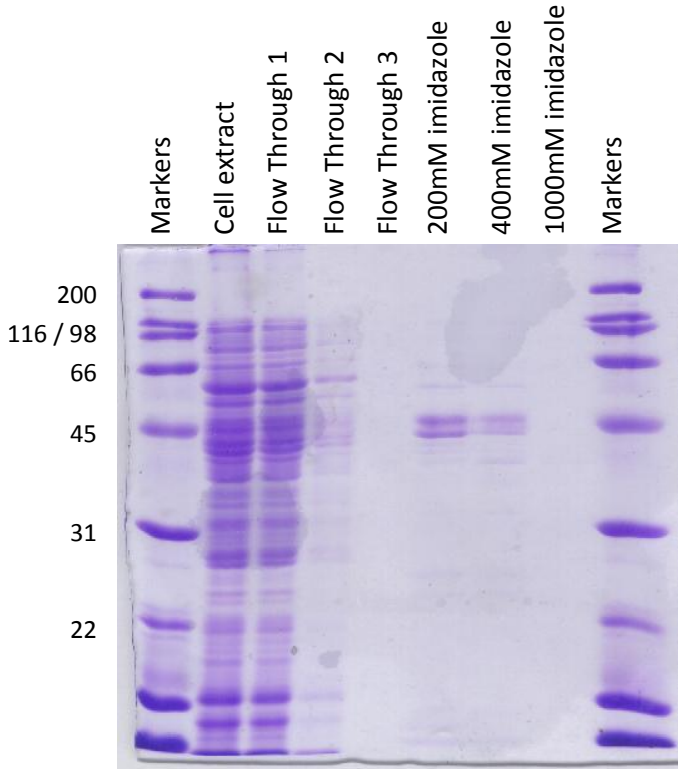


Fig. 4.13. SDS PAGE gel visualised with Coomassie blue stain. A band of the expected size for the recombinant xylonate dehydratase with both His and Strep tags (~48kDa) can be seen in the 20% and 40% elution fractions. The sizes of the Broad-Range markers (Biorad) are shown to the left of the figure in kDa.

To assess the effect of imidazole on the enzyme assay, a semi-purified native xylonate dehydratase (used as a positive control in enzyme assays) was incubated in the presence of 200mM or 400mM imidazole for 5 min and the remaining activity determined. In both

cases, activity was found to be significantly reduced compared to a control incubated with no imidazole (Fig. 4.14). Whether this decrease in activity is due to inhibition of the enzyme or a reaction with the assay is, however, unclear.

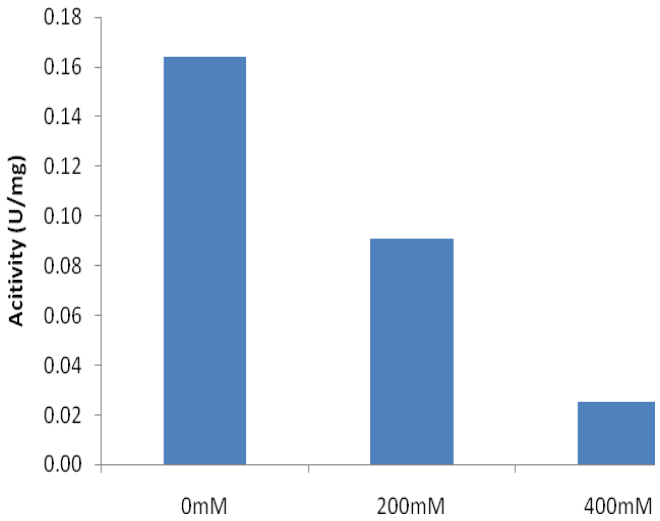


Fig. 4.14. Effect of imidazole on a semi purified native enzyme assayed under standard condition at 70°C as described in general Materials and Methods. One unit of enzyme activity is defined as the amount of enzyme required to form of 1μmol of 2-keto-3-deoxypentulosonic acid per min.

The nickel affinity purification was repeated with a fresh cell extract made up in 50mM Tris, pH8.0 containing 20mM $MgCl_2$ and 5mM imidazole to prevent non-specific binding. After loading the sample and washing the column with six volumes of this buffer, the bound material was eluted with 50mM imidazole (Fig. 4.15), a four-fold lower concentration than previously used. The reduction in imidazole resulted in detectable dehydratase activity in this purified fraction. In addition to the 48kDa recombinant protein (containing both His and Strep tags), a second slightly smaller band of equal intensity could also be seen when this fraction was visualised on a SDS PAGE gel. This smaller band is likely to be the native xylonate dehydratase associating and co-purifying with the recombinant version of the protein.

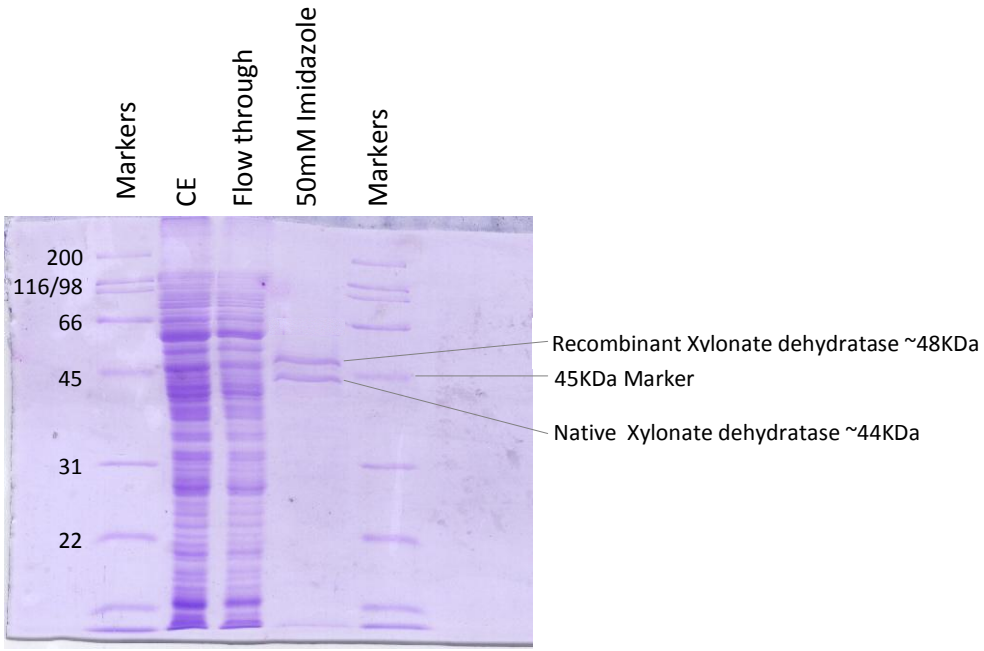


Fig. 4.15. SDS PAGE gel visualised with Coomassie Blue stain. A band of the expected size for the recombinant, tagged xylonate dehydratase (~48kDa) and native xylonate dehydratase (~44kDa) can be seen in the 50mM elution fraction. The sizes of the Broad-Range markers (Biorad) are shown to the left of the figure in kDa.

Fractions were assayed as before and the specific activity of xylonate dehydratase was found to be ten-fold higher in the purified sample compared to the unfractionated cell extract and flow through fractions from purification (Table 4.03, Fig. 4.16).

Sample	Activity (U/mg)
Cell Extract	0.019
Flow Through 1	0.016
Flow Through 2	0.000
50mM elution	0.251
250mM elution	0.015
500mM elution	0.000

Table 4.03. Xylonate dehydratase activity in fractions from a nickel affinity purification of D-arabinose induced cells. One unit (U) of enzyme activity is defined as the amount of enzyme required to form of 1μmol 2-keto-3-deoxypentulosonic acid per minute.

The native xylonate dehydratase enzyme has been shown to have a lower activity with L-arabinonate and no detectable activity with D-gluconate (chapter 3). The 50mM elution fraction was tested with both these substrates and showed the same activity profile as the native dehydratase (Fig. 4.17), indicating that it is a specific C5 dehydratase.

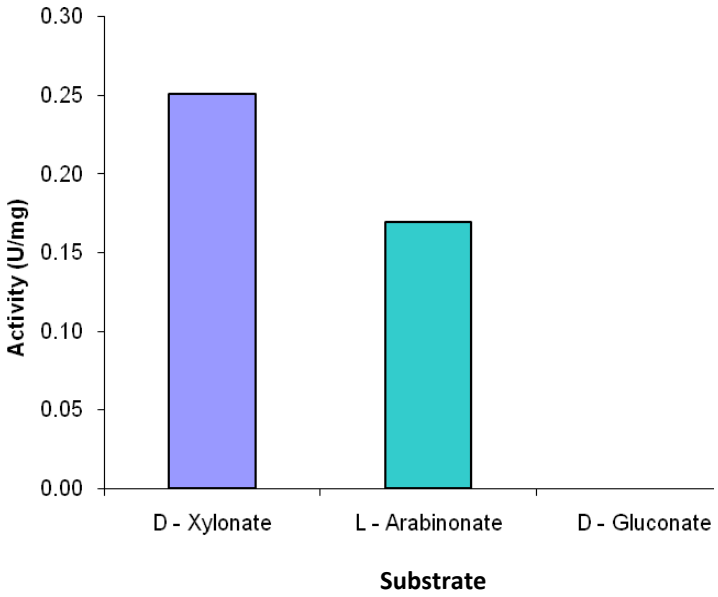


Fig. 4.17. Dehydratase activity in a 50mM imidazole elution fractions from a nickel affinity purification of D-arabinose induced cells. The fraction was assayed at 70°C with D-xlyonate, L-arabinonate or D-gluconate. One unit (U) of enzyme activity is defined as the amount of enzyme required to form of 1μmol 2-keto-3-deoxypentulosonic acid (xylonate and arabinonate substrates) or 2-keto-3-deoxygluconate (gluconate substrate) per min.

Unfortunately, due to the relatively low activity of the purified fraction, it was not possible to determine accurate kinetic data with D-xylonate and L-arabinonate as substrates. Errors associated with performing a stopped assay meant that activity was not significant or accurately reproducible under low substrate conditions.

4.3.2. Malate synthase

4.3.2.1. Gene cloning and heterologous expression

The putative malate synthase gene annotated in the genome of *S. solfataricus* (SSO1334) was PCR amplified producing a fragment of ~2.5kb (Fig. 4.18) as predicted by the genome sequence. Owing to the large size of the gene, it was cloned directly into an expression vector, where it was sequenced to confirm the fidelity of the PCR reaction. The putative malate synthase gene was expressed as a soluble, active recombinant protein in *E. coli.* The protein was purified to homogeneity by heat treatment, anion exchange and size exclusion chromatography as described. The molecular weight of the monomer was confirmed as 94kDa by SDS PAGE (Fig. 4.19) as predicted from the amino acid sequence. Gel filtration showed that the protein is active as a homodimer of M_r 188,000, like the homologous protein in *S. acidocaldarius*.

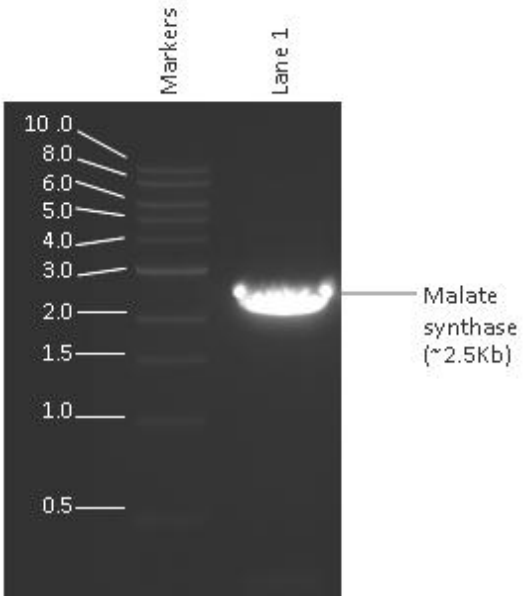


Fig. 4.18. 1% Agarose gel visualised with ethidium bromide under UV light. Lane 1 - PCR amplified malate synthase gene from *S. solfataricus*. The PCR product is approx 2.5kb as predicted by the gene sequence. Marker sizes are shown in kb on the left of the figure.

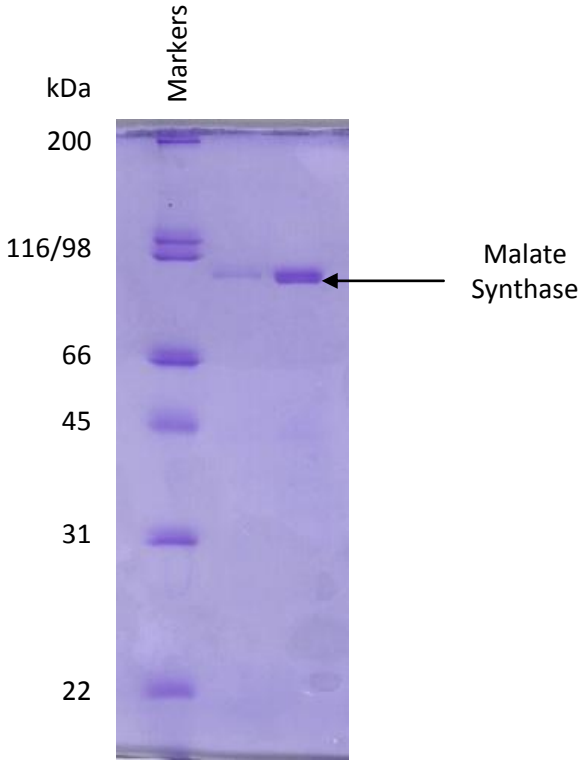


Fig. 4.19. SDS PAGE gel of the recombinant malate synthase purified to homogeneity by heat treatment, anion exchange and size exclusion chromatography. The monomeric protein is approximately 94kDa as predicted by the amino acid sequence. Markers are shown with their molecular weight in kDa on the left of the figure.

4.3.2.2. Kinetic Analysis

Kinetic parameters were determined for the recombinant enzyme, as described in Materials and Methods, for both glyoxylate and acetyl Co-A (Fig. 4.20 and Table 4.04).

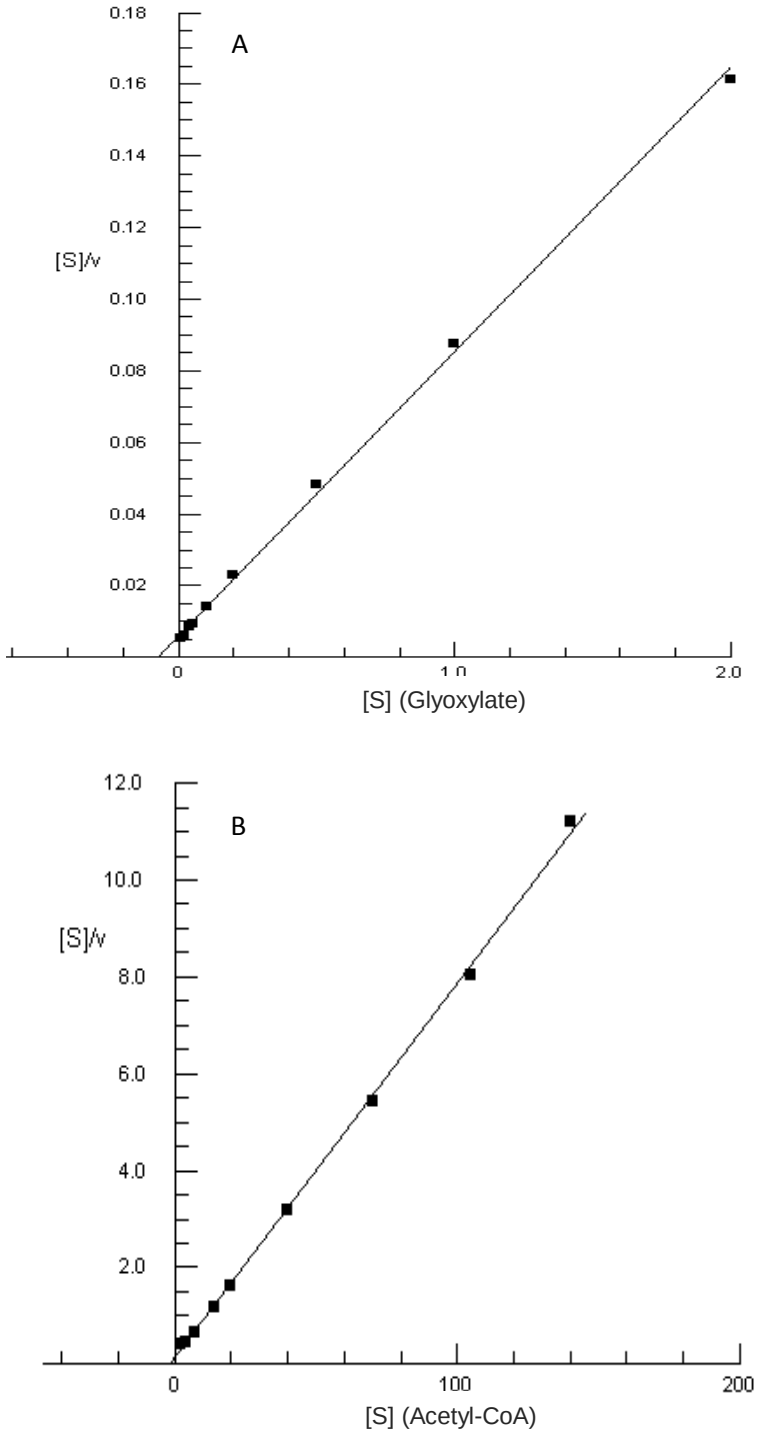


Fig. 4.20. Hanes-Woolf plots of malate synthase at 70°C with A) 0-2mM glyoxylate and 0.14mM Acetyl-CoA where [S] is shown in mM and B) 0-140μM Acetyl-CoA and 10mM glyoxylate where [S] is shown in μM. In both cases v is U/ml where one unit of enzyme activity is defined as the amount of enzyme required to produce 1μmol of 5-mercapto-2-nitrobenzoic acid per min.

Substrate	K_M (µM)	V_{max} (Umg)
Glyoxylate	61 (±2.0)	12.6 (±0.2)
Acetyl-CoA	2 (±0.1)	13.8 (±0.1)

Table 4.04. Kinetic parameters of recombinant malate synthase at 70°C. Values were determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

4.3.2.3. Thermoactivity and thermostability

Thermoactivity assays were performed as described in general Materials and Methods. Malate synthase showed maximum activity at 80°C under saturating substrate conditions (Fig. 4.21). This optimum is close to the optimal growth temperature of *S. solfataricus*. Thermal inactivation assays showed that the enzyme had half life of 5 min at 85°C (Fig. 4.22). However, correlation with the activity optimum needs a more detailed analysis (Chapter 6) as the inactivation experiment was carried out in the absence of substrates.

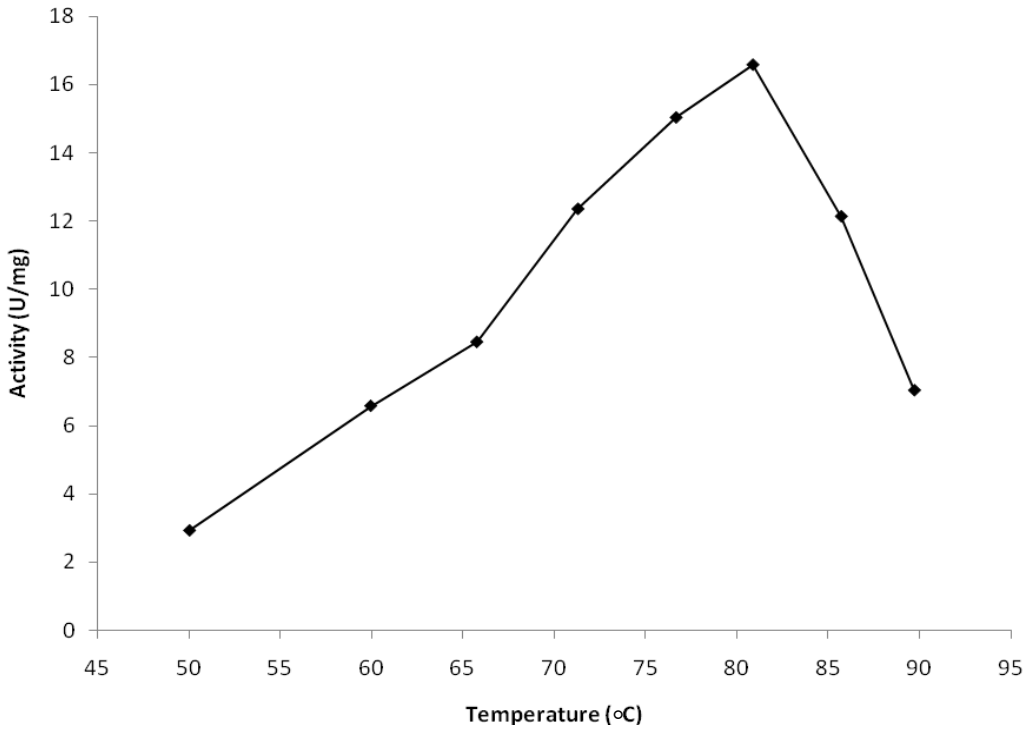


Fig. 4.21. Variation of malate synthase activity with increasing temperature measured under standard assay conditions. One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

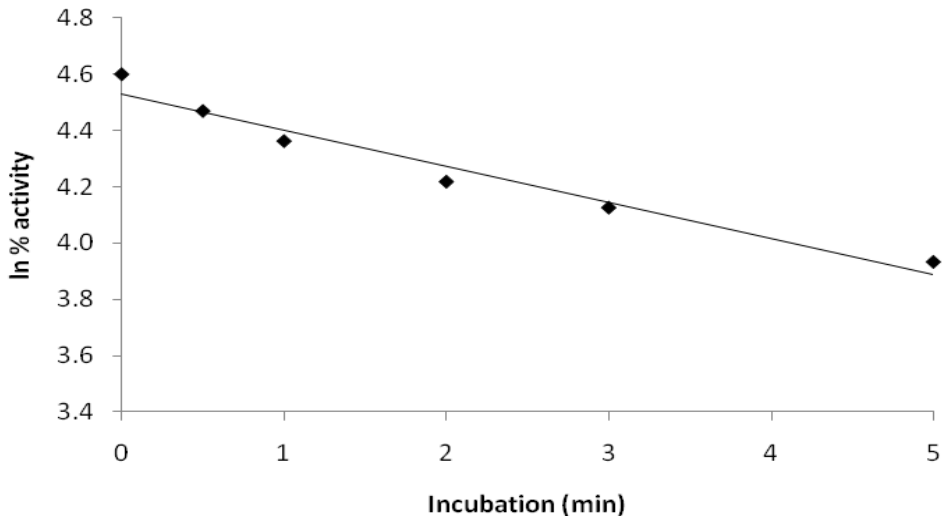


Fig. 4.22. Thermal inactivation of malate synthase at 85°C as described in general Materials and Methods.

4.3.2.4. Crystallisation Trials

Malate synthase was purified to homogeneity as described and concentrated to ~8mg/ml. Crystal screens JCSG+96 and Clear Strategy II (Molecular Dimensions, Suffolk UK) listed in Appendix 3 were set up using the sitting drop method with three protein concentrations in each condition (1:1, 2:1 and 3:1 protein to screen) in a total of 300 – 450nl. Trays were incubated at 15°C and checked every 24h for crystal growth.

Out of 192 conditions in the initial screen, only one condition (0.1M sodium cacodylate pH6.5, 40%(v/v) MPD, 5%(w/v) PEG 8000) produced crystals after incubation at 15°C for 3 days (Fig. 4.23).

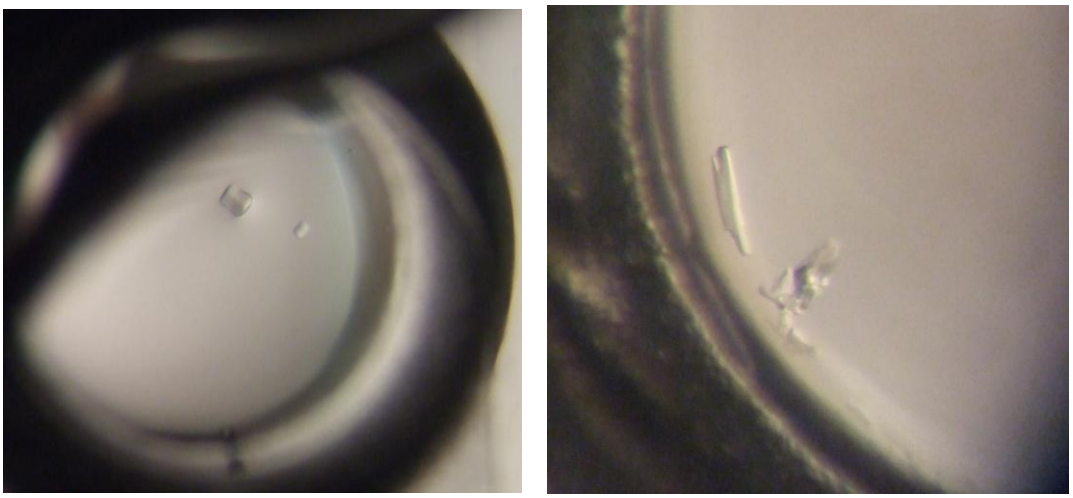


Fig. 4.23. Crystals obtained with recombinant malate synthase under sitting drop conditions in 0.1M sodium cacodylate pH6.5, 40%(v/v) MPD, 5%(w/v) PEG 8000.

This condition was comprehensively screened, varying the percentage of MPD and PEG and the concentration of protein, but larger crystals could not be obtained by either sitting drop or hanging drop methods.

4.3.3. Glycolate dehydrogenase (Glyoxylate reductase)

4.3.3.1. Gene cloning and heterologous expression

The putative glyoxylate reductase gene from *S. solfataricus* (SSO3187) was PCR-amplified (Fig. 4.24) as described and was ~0.95kb, as predicted from the genome sequence. The gene was cloned into the untagged vector pET3a, and the fidelity of the PCR reaction confirmed by sequencing.

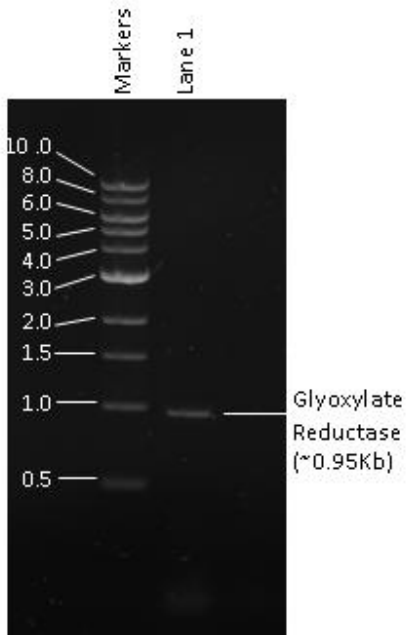


Fig. 4.24. 1% Agarose gel visualised with ethidium bromide under UV light. Lane 1 - PCR amplification of glyoxylate reductase gene from *S. solfataricus*. The PCR product is approx 0.95kb, as predicted by the gene sequence. Marker sizes are shown in kb on the left of the figure.

The untagged construct was expressed in Rosetta^{TM} as a soluble active recombinant protein, which was purified by heat treatment and size exclusion chromatography as described in Materials and Methods. The molecular weight of the recombinant monomer was confirmed as 34kDa by SDS PAGE (Fig. 4.25) as predicted from the amino acid sequence. Analysis by gel filtration showed that the enzyme is active as a homodimer of M_r 62,000.

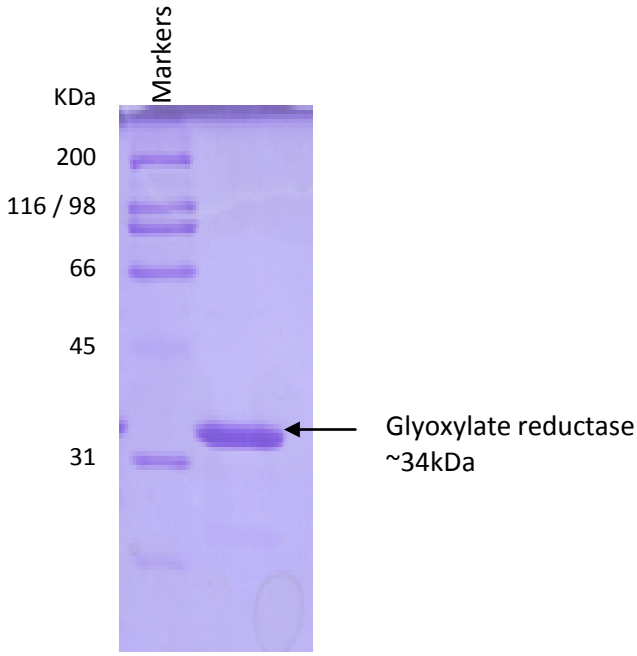


Fig. 4.25. SDS PAGE gel of the recombinant glyoxylate reductase, purified to homogeneity by heat treatment and size exclusion chromatography. The monomeric protein is approximately 34kDa as predicted by the amino acid sequence. Markers are shown with their molecular weight in kDa on the left of the figure.

4.3.3.2. Kinetic Analysis

The recombinant protein had high activity with glyoxylate and NADH (Table 4.05, Fig. 4.26), but owing to the high K_M for NADH, it was not possible to calculate kinetic parameters for glyoxylate under saturating concentrations of cofactor, as the absorbance would have been too high to collect accurate data. Instead, apparent values were calculated and the V_{max} adjusted by calculating the percentage of enzyme saturation at this lower substrate concentration using Equation 4.01.

Eqn. 4.01

$$\frac{v}{V_{max}} = \frac{[S]}{K_M + [S]}$$

Substrate	K_M (mM)	V_{max} (U/mg)	Adjusted V_{max} (U/mg)
Glyoxylate	4.7 (±0.10)	158 (±1)	193 (±1)
NADH	0.09 (±0.003)	169 (±2)	188 (±2)

Table 4.05. Kinetic parameters of recombinant glyoxylate reductase at 70°C determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974) and the adjusted values for V_{max} determined by calculating the % saturation of the enzyme using equation 4.01 (calculated as 90% saturation with glyoxylate and 82% with NADH). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

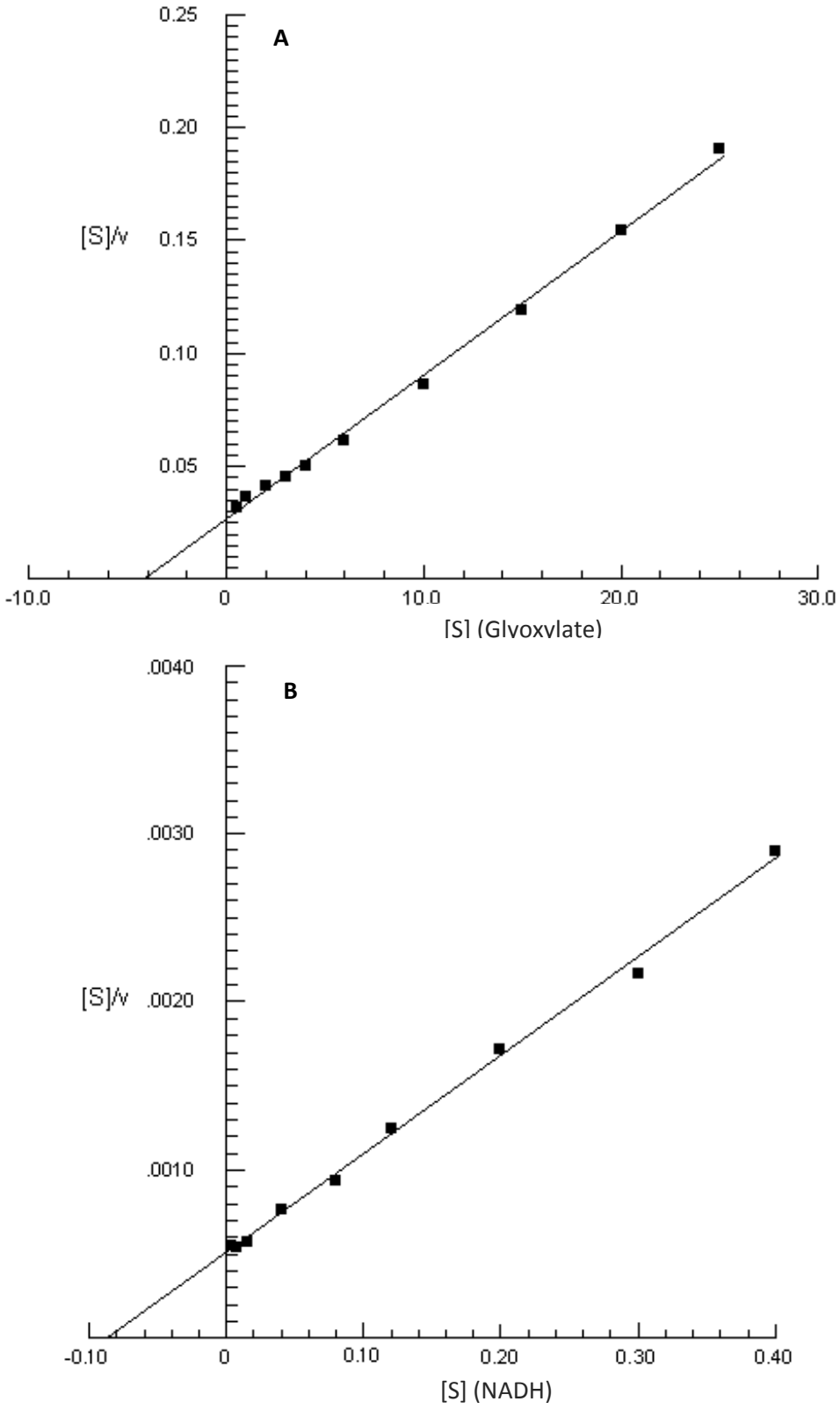


Fig. 4.26. Hanes - Woolf plots of recombinant glyoxylate reductase activity at 70°C with (A) 0-25mM glyoxylate and 0.4mM NADH or (B) 0-0.4mM NADH and 50mM glyoxylate. [S] is shown in mM and v in U/mg where one unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

The recombinant glyoxylate reductase also showed activity with NADPH as a cofactor, but there appeared to be slight substrate inhibition at high concentrations of glyoxylate (Fig. 4.27). Data were fitted to the substrate inhibition equation (pg. 21) and values determined for K_M = 0.68 (±0.05)mM, V_{max} = 46 (±2)U/mg, and K_{si} = 52 (±10)mM.

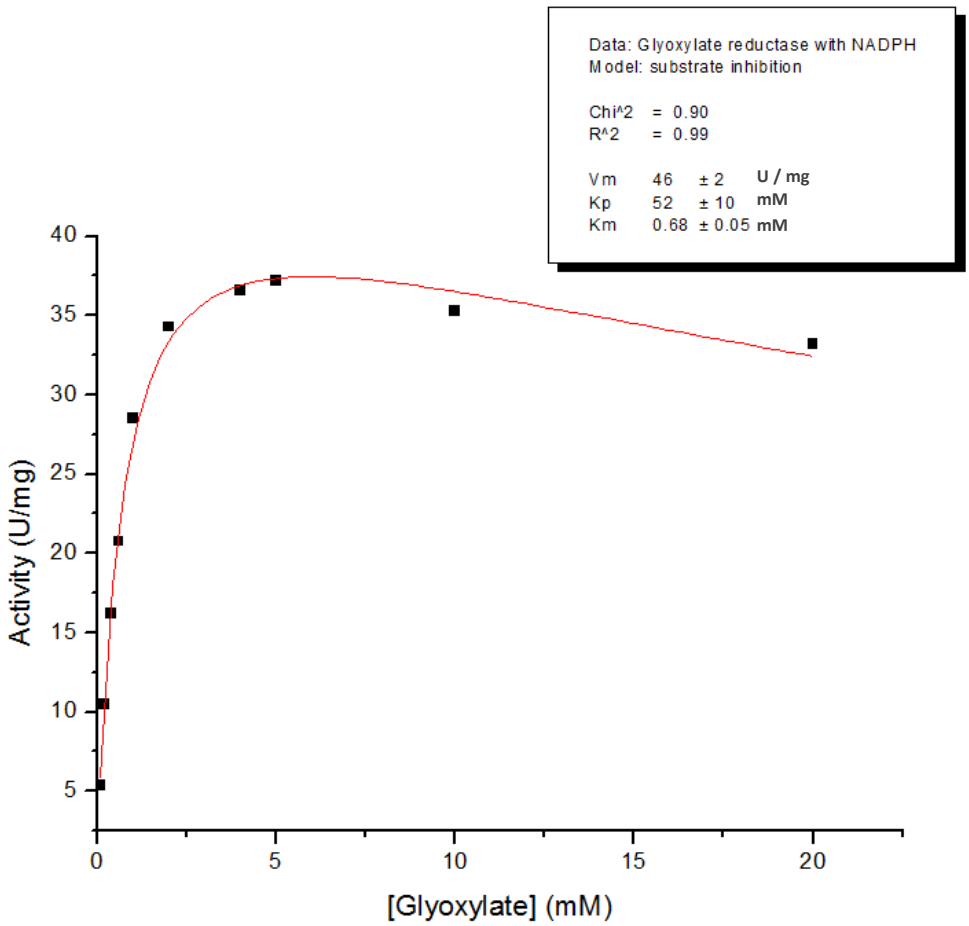


Fig. 4.27. Activity of recombinant glyoxylate reductase at 70°C with 0-20mM glyoxylate and 0.2mM NADPH as a cofactor. Data were fitted to the substrate inhibition equation (pg.21). One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute

It was not possible to calculate a true K_M and V_{max} for NADPH due to the inhibition at saturating concentrations of glyoxylate. Apparent values were, therefore, determined at 5mM glyoxylate and the V_{max} adjusted by calculating the percentage of enzyme saturation at this lower substrate concentration using Equation 4.01 (Table 4.06).

Substrate	K_M^{APP}	V_{max}^{APP}	Adjusted V_{max}
NADPH	9.0 (±0.5)μM	41 (±0.03)U/mg	47 (±0.03)U/mg

Table 4.06. Apparent kinetic parameters of recombinant glyoxylate reductase at 70°C determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974) and the adjusted value determined from the % saturation of the enzyme with glyoxylate (calculated as 88%). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1μmol of $NADP^+$ per minute.

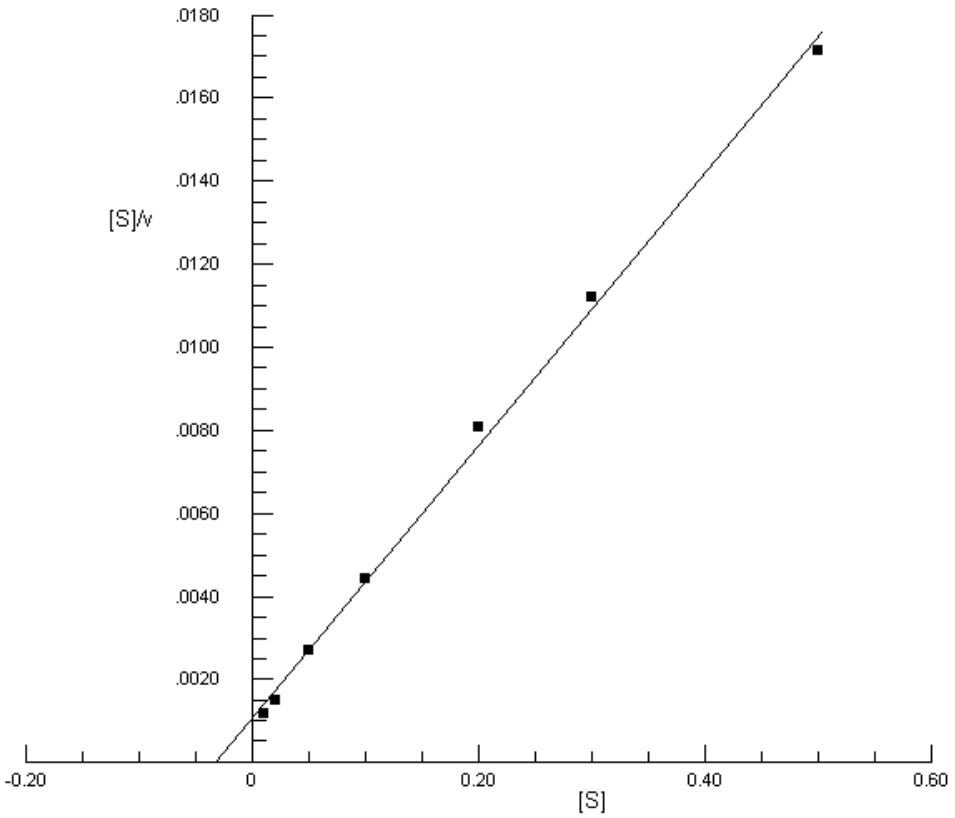


Fig. 4.28. Hanes - Woolf plot of recombinant glyoxylate reductase activity at 70°C with 50mM glyoxylate and 0 - 0.5mM NADPH. [S] is shown in mM and v in U/mg where one unit of enzyme activity is defined as the amount of enzyme required to produce 1μmol of $NADP^+$ per minute.

4.3.3.3. Thermoactivity and thermostability

Glyoxylate reductase was found to be thermostable with a t½ of 35min at 85°C (Fig. 4.30). Determination of initial rates at different temperatures showed maximum activity at 70°C (Fig. 4.29), 10°C lower than the optimum growth temperature of *S. solfataricus.*

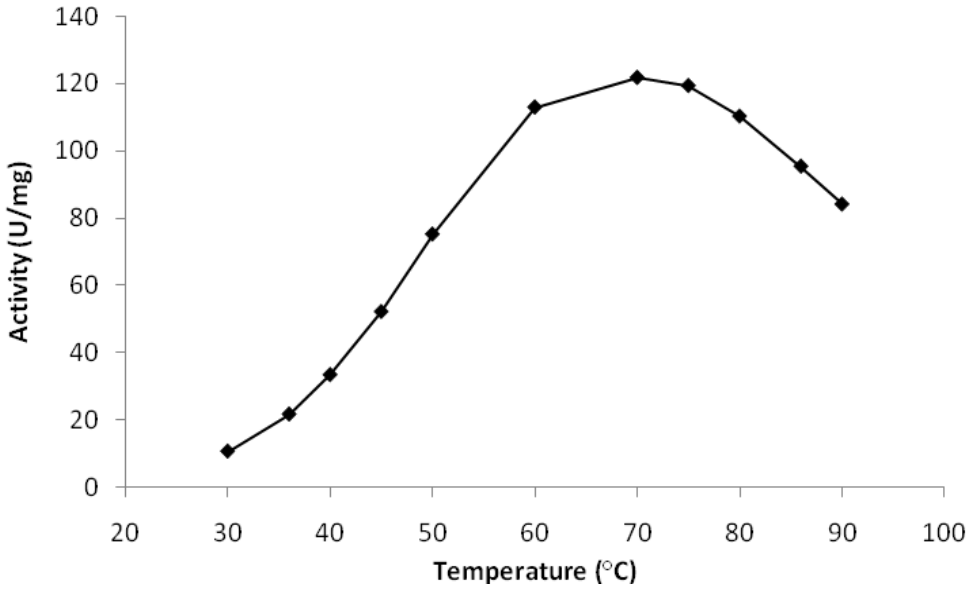


Fig. 4.29. Variation of glyoxylate reductase activity with increasing temperature with 50mM Glyoxylate and 0.4mM NADH.

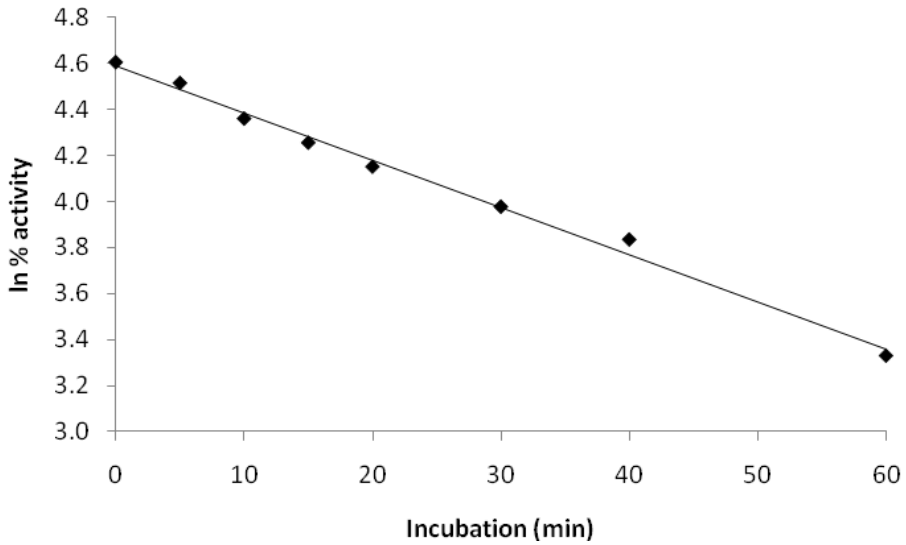


Fig. 4.30. Thermal inactivation of glyoxylate reductase at 85°C as described in general Materials and Methods.

4.3.3.4. Oxidation of glycolate

A coupled assay to detect the oxidation of glycolate (Fig. 4.31) was performed as described in Materials and Methods, utilising the essentially irreversible malate synthase reaction to drive the equilibrium towards the formation of glyoxylate. Due to a significant background rate under standard assay conditions used previously at pH7.5 and pH8, the assay was performed in 50mM sodium phosphate buffer at pH6.0, which reduced the background rate. The assay mixture contained an excess of NAD^+ (2mM) and acetyl-CoA (0.14mM), 10mM glycolate and 20µg pure recombinant malate synthase enzyme. The reaction was started by the addition of glyoxylate reductase and the rate of NADH production was determined to be 0.32 (±0.02) U/ml.

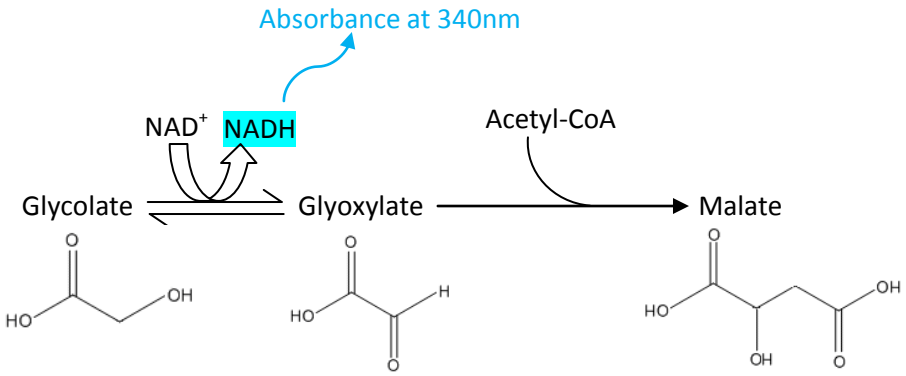


Fig. 4.31. Coupled assay to detect the oxidation of glycolate by the enzyme glyoxylate reductase; the essentially irreversible condensation of glyoxylate and acetyl-CoA pulls the equilibrium towards the formation of glyoxylate.

4.3.3.5. Substrate promiscuity

The recombinant glyoxylate reductase was assayed with a range of substrates to determine whether it showed the same substrate promiscuity identified in glyoxylate reductase enzymes from other organisms (Fauvart *et al*, 2007; Yoshikawa *et al*, 2007; Ohshima *et al*, 2001). Owing to the apparent inhibition found with NADPH, NADH was used as a co-factor when determining kinetic parameters (Table 4.07 and Fig. 4.32).

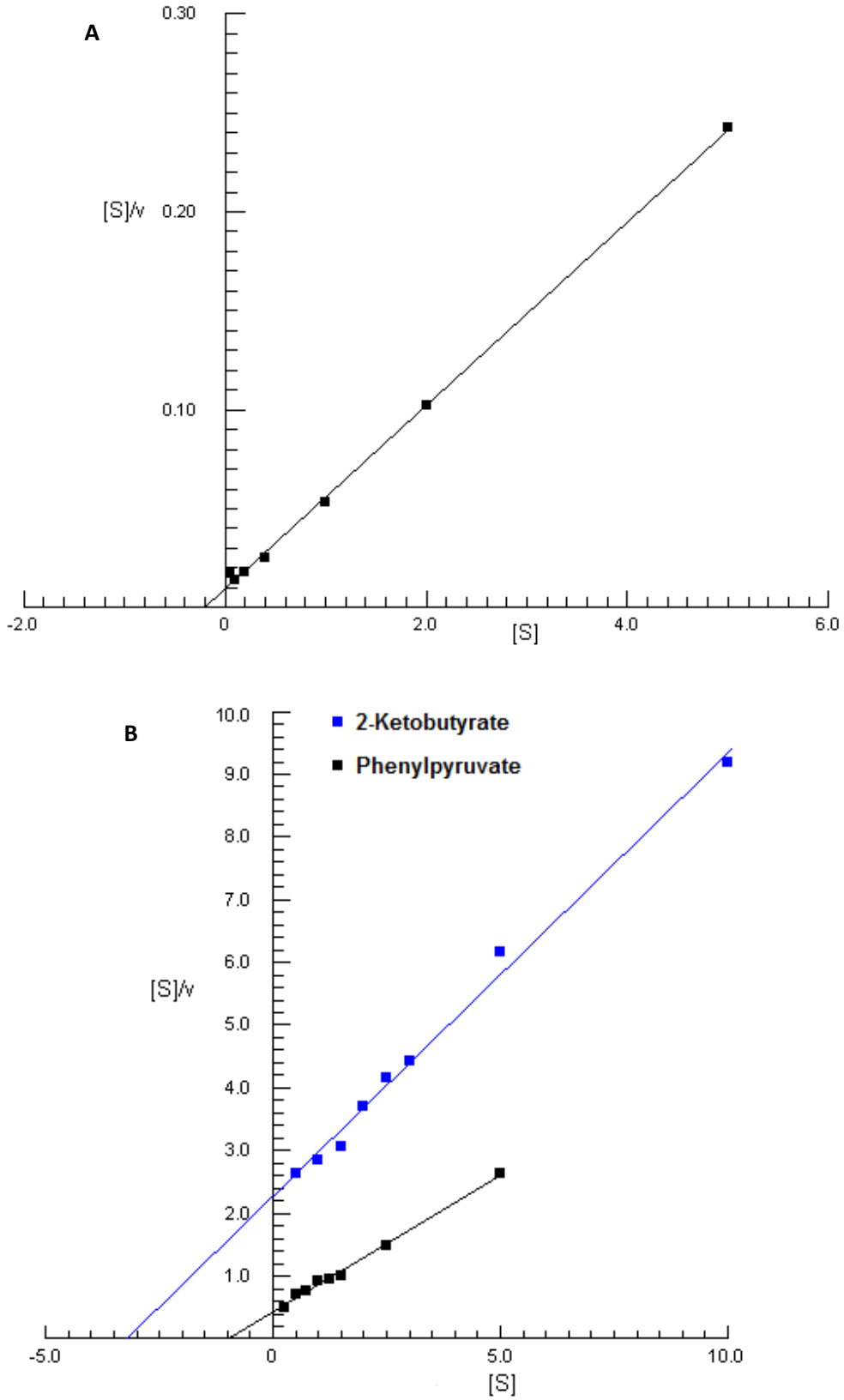


Fig. 4.32. Hanes – Woolf plots of recombinant glyoxylate reductase activity with (A) hydroxypyruvate and (B) 2-ketobutyrate or phenylpyruvate as substrates at 70°C with an NADH as a cofactor (0.4mM). [S] is shown in mM and v in U/mg where one unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

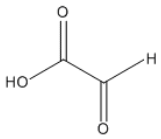
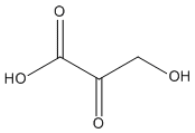
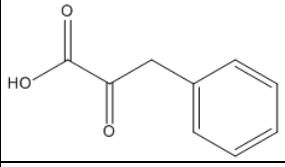
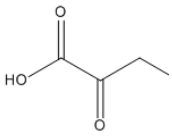
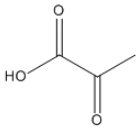
Substrate	Structure	K_M (mM)	V_{max} (U/mg)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
Glyoxylate		4.7 (±0.10)	193 (±2)	23
Hydroxypyruvate		0.2 (±0.01)	23 (±1)	62
Phenylpyruvate		1.1 (±0.04)	2 (±0.03)	1.2
2-ketobutyrate		2.9 (±0.15)	1 (±0.03)	0.3
Pyruvate		-	-	

Table 4.07. Comparison of structure and kinetic parameters with glyoxylate, hydroxypyruvate, phenylpyruvate and 2-ketobutyrate as substrates at 70°C with an NADH as a cofactor (0.4mM). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NAD^+ per minute.

The V_{max} values for alternative substrates were all lower than for primary substrate glyoxylate. However, the K_M for hydroxypyruvate was also significantly lower than that of glyoxylate, resulting in a higher k_{cat}/K_M for this substrate. Reduction of 2-oxoisocaproate was also detected, but the activity was too low to establish accurate kinetic parameters for this substrate. Interestingly, no activity was observed with pyruvate, despite its structural similarity with glyoxylate and 2-ketobutyrate.

4.4. Discussion

4.4.1. D-Xylonate dehydratase

The use of *E. coli* strains as a host for recombinant protein production was the most appropriate starting point for expression of the recombinant xylonate dehydratase, given the previous success with other recombinant proteins cloned from *S. solfataricus*. The first expression trials were carried out with BL21(DE3) using different induction methods (IPTG induced, uninduced and autoinduced using Overnight Express™ media) which proved to be unsuccessful.

The same conditions were repeated using Rosetta™ competent cells, which were selected as they contain the pRARE2 plasmid, enabling the host to efficiently express genes containing the rare codons frequently found in archaeal genes. After completing initial expression trials at 37°C throughout, a second series of expression trials was attempted with Rosetta™, this time increasing the temperature to 45°C after induction as described by Koma *et al* (2006), who used this technique to successfully increase the yield of thermophilic proteins in *E coli*. Unfortunately, in the case of xylonate dehydratase, the change in temperature made no difference to the expression level or solubility of the protein.

The apparent lack of expression of the untagged construct suggested that there may be toxicity associated with the active protein, resulting in its degradation by the host cell. This prompted the use of the BL21(DE3) mutant strains C41 and C43, which have been shown to efficiently express a range of potentially toxic proteins (Dumon-Seignovert *et al*, 2003), but again no soluble protein or enzyme activity could be detected under these conditions

Work in this study and others (Kim & Lee, 2005) has indicated that the xylonate dehydratase and gluconate dehydratase from *S. solfataricus* are phosphorylated for activity, which may not be achieved in a bacterial expression system. This information coupled with previous studies indicating that only 50% of full length archaeal proteins can be expressed as soluble proteins in *E. coli* (Graslund *et al*, 2007) led to the utilisation of a recently developed homologous *Sulfolobus* expression system (Albers *et al*, 2006).

Using this system, it was possible to produce a soluble recombinant xylonate dehydratase. The protein was purified to near homogeneity using nickel affinity chromatography and was the expected size of 48kDa (due to the presence of His and Strep tags) when analysed on SDS PAGE. Interestingly, a second band of equal intensity was also observed which corresponded in size to the native xylonate dehydratase, which was approximately 44kDa. It is possible that due to the oligomeric nature of the native protein, (as described in chapter 2) an association between the native and recombinant polypeptides has occurred, leading to this co-purification.

The activity observed in this purified fraction was comparatively low and did not allow the determination of kinetic parameters. It is possible that only a percentage of the protein present was in a phosphorylated and active form or that the activity being measured is actually due to the associated native protein which appears to co-purify. However, it is important to state that the *Sulfolobus* system is not expected to produce the same level of overexpression that is observed for recombinant proteins in bacterial systems. This is due in part to the comparatively low growth rate of *Sulfolobus* compared to *E. coli*, the former producing approximately 1g of cell paste per litre of culture and often reaching a maximum OD_{600} of just 0.6, over tenfold lower than that expected for bacterial systems.

While the results from these expression trials are not overwhelming, they do provide some evidence that the putative xylonate dehydratase identified is responsible for the observed activity in *S. solfataricus*. There are numerous examples of studies (Gerlt *et al*, 2005; Rakus *et al*, 2008; Rakus, *et al,* 2009) that have concluded members of the mandelate racemase family have evolved through gene duplications and evolution – often from promiscuous enzymes. The closely related organism *S. acidocaldarius* possesses a single gluconate dehydratase enzyme, capable of catalysing the dehydration of both C5 and C6 substrates (P Schönheit P, University of Kiel, Germany, personal communication). It is therefore extremely likely that in *S. solfataricus*, gluconate and xylonate dehydratase have evolved from a single promiscuous enzyme through an initial gene duplication event.

4.4.2. Malate synthase

The putative malate synthase gene from *S. solfataricus* was cloned and expressed as a soluble active protein and represents the first recombinant malate synthase produced from a thermophilic organism. The kinetic parameters determined for the recombinant malate

synthase concur with those values found in a native cell extract, indicating that the correct enzyme has been cloned and expressed. Like the homologous enzyme identified in *S. acidocaldarius*, the *S. solfataricus* enzyme is functional as a homodimer and does not require Mg^{2+} for activity.

Two isoforms of malate synthase have previously been identified; isoform A which is found in fungi and plants, and isoform G which is found only in bacteria and is larger by approximately 200 residues (Lohman *et al,* 2008); however, neither of these forms has been reported to exist in Archaea. Malate synthase from the halophilic archaeon *Haloarcula marismortui* has a significantly smaller subunit size (~100 residues less) than isoform A, the same as reported in *Haloferax volcanii* (Serrano *et al*, 2001). In contrast to this, malate synthase enzymes from the thermophilic archaea a *Pyrococcus*, *Sulfolobus* and *Thermoproteus tenax* appear to be larger than isoform G by approximately 100 residues. Malate synthase enzymes from *Pyrococcus* and *Thermoproteus* have not yet been characterised, but their sequence similarity with the *Sulfolobus* sp. enzyme (see Appendix 4) suggests that they may form a distinct class of thermophilic malate synthases, with modified reaction mechanism that does not require Mg^{2+} for activity. Crystallisation trials were set up to provide structural information to investigate the mechanism of substrate binding, but attempts to produce large enough crystals to gain this information were unsuccessful.

The enzyme from *S. solfataricus* is moderately thermostable displaying maximum activity at approximately 80°C, but loses 30% of activity when the temperature is increased to 85°C. In the classical or 'two state' model, loss of activity with temperature is assumed to be through irreversible thermal denaturation. However, as the activity at each temperature was calculated from the first 20 seconds of the enzyme reaction, this dramatic reduction in activity over the 5°C temperature change cannot be accounted for by irreversible enzyme denaturation, which occurs over a period of several minutes rather than seconds (Daniel *et al* 2001). This effect has been reported previously, most notably by Thomas and Scopes (1998) who studied the effect of temperature on mesophilic and thermophilic 3-phosphoglycerate kinases. Anomalies such as these have challenged the existing two state model and lead to the development of a new model to describe the effect of temperature on enzyme activity. Further investigation into the thermoactivity and thermostability of malate synthase, with respect to this new model is presented in chapter 6.

4.4.3. Glycolate dehydrogenase (Glyoxylate Reductase)

The recombinant glyoxylate reductase was heterologously expressed as a soluble, active enzyme in *E. coli*. The K_M for glyoxylate was similar to that previously reported in a native cell extract of *S. solfataricus* (Chapter 3), indicating that the correct gene product had been identified. Owing to the equilibrium position of the reaction *in vitro,* glyoxylate reductase was assayed in the opposite direction to which it is proposed to operate in C5 catabolism, where the action of subsequent enzymes in this metabolic pathway would draw the equilibrium towards the formation of glyoxylate. By coupling the recombinant glyoxylate reductase and malate synthase it was possible to demonstrate that the reaction would also proceed in the 'forwards' direction generating glyoxylate from glycolate.

As expected from thermophilic proteins, the recombinant enzyme was thermostable with a half life of 35min at 85°C, but showed maximum activity at 70°C, over 10°C lower than the optimum growth temperature of *S. solfataricus*. However, due to the different conditions under which these data were collected (either in the presence or absence of substrates) which may have affected the stability of the protein, the relationship between the activity of the enzyme at a specific temperature and its thermostability is difficult to ascertain. As K_M has been reported to increase with temperature (Scopes, 1995), glyoxylate reductase would not be a suitable enzyme for further thermoactivity and thermostability studies, due to the relatively high K_M for both glyoxylate and NADH, which would prevent it from being saturated at higher temperatures.

Like glyoxylate reductase enzymes from other archaea (Yoshikawa *et al*, 2007; Ohshima *et al,* 2001), glyoxylate reductase from *S. solfataricus* can utilise both NADH and NADPH as a cofactor. Structural studies on glyoxylate reductase from *P. horikoshii* identified an arginine residue (R181) crucial for recognition of the 2'-phosphate group of NADPH (Yoshikawa *et al*, 2007) and an alignment of archaeal glyoxylate reductase sequences revealed that this residue is conserved in *S. solfataricus*. Interestingly, although product inhibition was observed with NADPH, the presence of the phosphate appears to have an effect on substrate binding, resulting in a decrease in K_M for glyoxylate. This lowering of K_M was also reported in glyoxylate reductase from the thermophilic bacterium *Thermus thermophilus* (Ogino *et al*, 2008) but not in other archaeal glyoxylate reductase enzymes.

Glyoxylate reductase from *S. solfataricus* was shown to have activity with a variety of substrates, although with a significantly lower rate than for the primary substrate, glyoxylate. Unlike other members of the 2-hydroxyacid dehydrogenase family, which appear to be specific for a single substrate, all bacterial and archaeal glyoxylate reductase enzymes studied so far have shown promiscuity towards glyoxylate and hydroxypyruvate (Ogino *et al*, 2008). Like glyoxylate reductase enzymes from other organisms, the enzyme from *S. solfataricus* catalyses the reduction of glyoxylate, hydroxypyruvate and phenylpruvate with different efficiencies. The only reported exception is glyoxylate reductase from *Rhizobium etli*, which reduces phenylpruvate with an equal efficiency to glyoxylate. The presence of substrate promiscuity has been associated with evolution of enzyme function (Aharoni *et al*, 2005), in view of the survival advantage it confers on organisms in a variable environments. While this conserved promiscuity is an interesting feature of glyoxylate reductase enzymes, the biological advantage of this promiscuity to the organism remains unclear (Fauvart *et al*, 2007).

Chapter 5: Investigation of substrate specificity through mutagenesis studies of glucose dehydrogenase from *S. solfataricus*

5.1. Introduction

Glucose dehydrogenase (GDH) catalyses the first step of the non-phosphorylative and part-phosphorylative variants of the Entner-Doudoroff pathway in *Sulfolobus solfataricus*. The recombinant enzyme has previously been characterised in terms of its kinetic properties and the structure has been determined to 1.5Å and 1.6Å complexed with glucose or xylose respectively (Milburn *et al,* 2006). It is a homotetrameric protein with a monomer size of 41kDa, and belongs to the medium chain dehydrogenase / reductase (MDR) superfamily. (Milburn *et al,* 2006). These enzymes are characterized by a chain length of 350 – 375 residues and conserved structural zinc-binding and nucleotide-binding sites (Edwards *et al* 1996).

GDH catalyses the oxidation of D-glucose, D-galactose, D-xylose and L-arabinose to their respective acids using either NAD^+ or $NADP^+$ as a cofactor, and contains the characteristic GXGXXG nucleotide-binding motif at residues 188 –193 (Lamble *et al* 2003; Milburn *et al,* 2006). Activity with other sugars such as glucosamine and fucose has also been reported (Lamble *et al*, 2003) but so far no kinetic parameters have been established for these substrates.

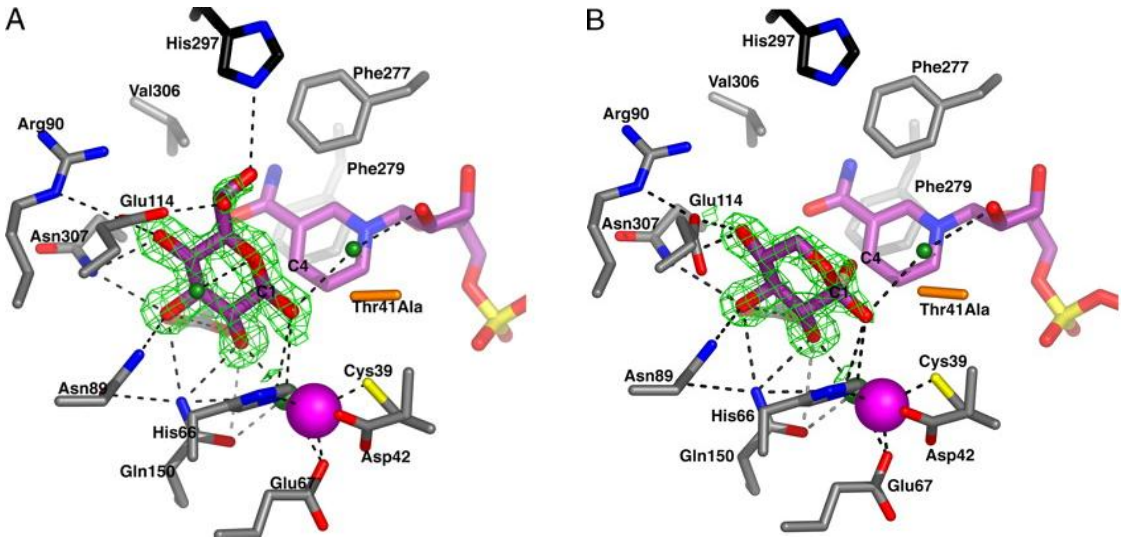


Fig. 5.01. Crystal structure of the GDH active site showing the hydrogen bonding between active site residues and (A) glucose and (B) xylose. The position of the $NADP^+$ cofactor and Zn^{2+} ion (pink sphere) are also shown. (Taken from Milburn *et al*, 2006).

The crystal structure of GDH shows that hydrogen bonding between key active site residues around the C2 and C3 hydroxyl groups of the substrate are critical for activity as they allow the C1 hydroxyl of the sugar to be close enough to the reactive C4 position of the nicotinamide ring of the co-factor to allow hydride transfer (Fig. 5.01) (Milburn *et al,* 2006). GDH assembles with H297 from one monomer contributing to the active site of the adjacent monomer, restricting the space at one end of the binding cleft (Milburn *et al*, 2006). This residue facilitates the positioning of glucose by creating a hydrogen bond at the C6 position, but does not participate in the interaction with xylose, where the C6 group is absent. The structures of natural GDH substrates all have a glucose specific stereo configuration at the C2 and C3 position but can have either glucose or galactose configuration at C4 (Fig. 5.02). Sugars that show the alternative stereo configuration at the C2 or C3 position (Fig. 5.03) are not oxidised by GDH as steric hindrance prevents these substrates being correctly positioned to allow hydride transfer at the C1 hydroxyl group.

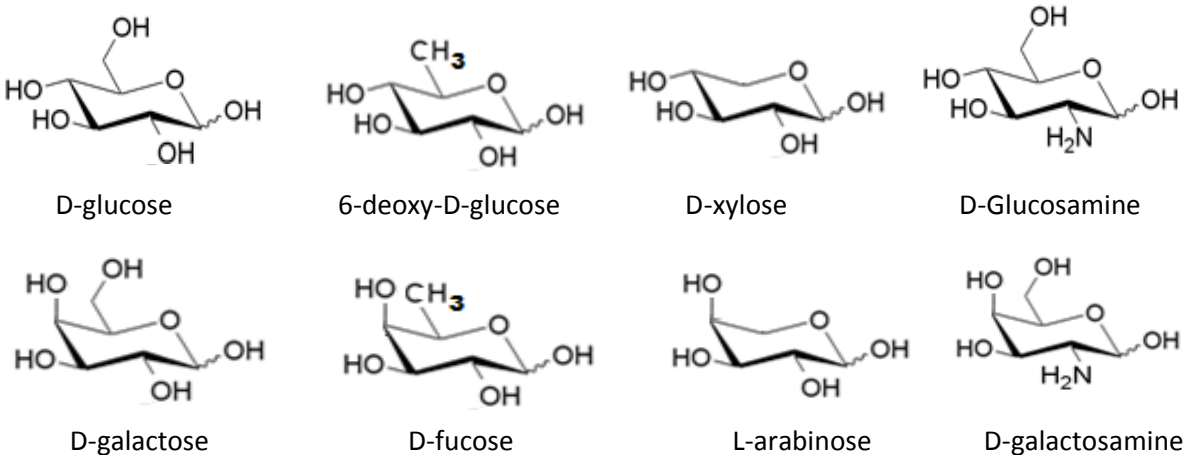


Fig. 5.02. Structures of known GDH substrates. All have glucose specific stereo configuration at the C2 and C3 position but can have either the glucose or galactose configuration at C4.

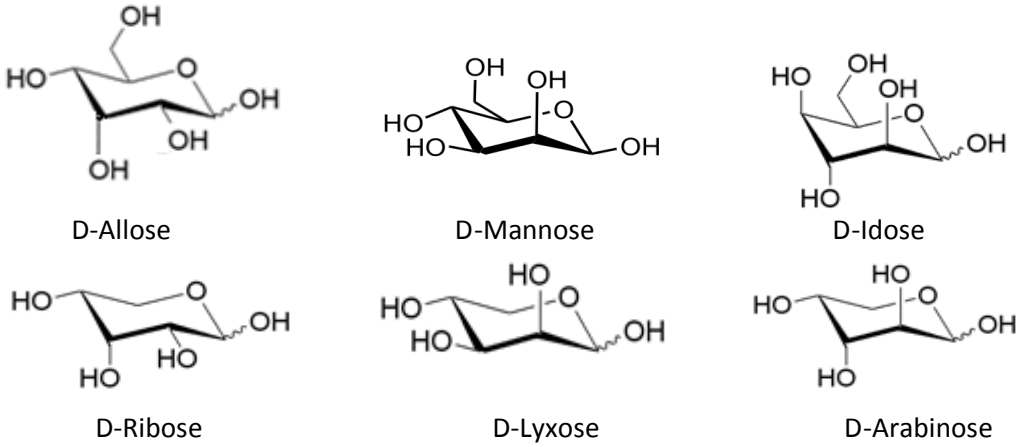


Fig. 5.03. Structures of sugars with which GDH shows no activity. All have alternative stereo configuration to glucose at the C2 or C3 positions, preventing correct positioning in the active site to allow hydride transfer at the C1 position.

The structure of GDH identifies nine hydrogen bonds between three key active site resides, Asp89, Gln150 and Asp154, and the C2 and C3 positions of glucose / galactose. To accommodate the different stereo configurations of L-xylose, D-mannose, D-Lyxose, D-idose, D-arabinose, D-allose and D-ribose, residues will be introduced which should result in hydrogen bonding at the C2 and C3 positions of these sugars, allowing for oxidation to take place.

Mutation of Asn89 to Asp might create an additional hydrogen bond with the C2 OH group of the epimer D-mannose. At the pH of the assay (pH8.0), the Asp carboxyl is deprotonated making it an acceptor for the proton of the sugar C2 hydroxyl, which should enhance the binding of the sugar in the active site and therefore allow it to be oxidised. An alternate mutation of Asn89 to Glu should allow bonding to occur at both the C2 and C3 positions of the D-mannose as the carboxyl group of Glu is deprotonated at pH8.0 and will therefore act as an acceptor to strengthen the bonding at these positions. A similar effect is predicted with the mutation Gln150 to Glu. Mutation of Asp154 to Leu should help to accommodate C3 epimers of glucose into the active site of the enzyme as the hydroxyl group in this position would be facing the $NADP^+$. By replacing Asp154 with Leu, the interaction with the side chains would no longer be limited.

The effect of introducing mutations on the stability of the enzyme must be considered in conjunction with the effect on function, as it has been widely recognised that mutations aimed at introducing new or enhancing existing promiscuous activities often have unfavourable effects on enzyme stability (Bloom *et al*, 2006; Bloom & Arnold, 2009). It has been reported that approximately 50% of mutations that confer new enzymatic activities result in a decrease in thermodynamic stability (Tokuriki *et al*, 2008). Residues in the active site region are not optimised for stability as they have functional constraints such as maintaining buried charges resulting in a 'trade off' between function and stability when they are mutated (Bloom *et al*, 2006). Further investigation of the connections between mutational effects, function and stability are therefore vital to developing a greater understanding of this relationship (Tokuriki & Tawfik, 2009).

5.2. Materials and Methods

5.2.1. Mutagenesis of glucose dehydrogenase

Mutagenesis reactions were carried out using a QuickChange II-E Site Directed Mutagenesis kit (Stratagene, West Lothian, UK) following the manufacturer's protocol.

Mutant	Codon change	Forward primer	T_m (°C)
N89D	aat → gat	ctggtaatgccagttgataggagaggatgcgg	79.9
N89E	aat → gag	ggagatctggtaatgccagttgagaggagaggatgcgg	83.5
N89A	aat → gct	ggagatctggtaatgccagttgctaggagaggatgcgg	83.5
Q150E	caa → gaa	ggaattttagcagaaccgttagcagacattgagaagtcc	80.8
D154L	gac → ctc	gcacaaccgttagcactcattgagaagtccattgagg	80.1
D154I	gac → atc	gcacaaccgttagcaatcattgagaagtccattgagg	80.1
D154N	gac → aat	gcacaaccgttagcaaatattgagaagtccattgagg	77.9
N307L	aat → ctt	cgataataggtctagtgcttggtcaaaagccccacttcc	80.5
H297F	cac → ttc	gacacaagaaatagtattcacaaataagacgataataggtc	75.8
H297Y	cac → tgg	gacacaagaaatagtatggacaaataagacgataataggtc	75.5
H297W	cac → tac	gacacaagaaatagtatacacaaataagacgataataggtc	77.0

Table 5.01. Forward primers designed for mutagenesis of the *S. solfataricus* wild type (WT) glucose dehydrogenase in pET3a. Reverse primers are direct reverse complement sequences and so are not shown. The melting temp (T_m) for each pair of primers was calculated using the formula; $T_m = 81.5 + 0.41$ (%GC) – 675/N - % Mismatch (N = Primer length in bases).

5.3. Results

5.3.1. Activity of glucose dehydrogenase mutants

Mutagenesis reactions were successfully carried out on the wild type (WT) glucose dehydrogenase (GDH) in pET-3a using QuickChange-IIE site-directed mutagenesis kit (Stratagene). GDH mutants were expressed overnight in BL21(DE3) cells without induction. Cell extracts were prepared and assayed for GDH activity with glucose and $NADP^+$ using the standard assay as described. Active recombinant proteins were subsequently purified by a single heat-treatment step at 70°C for 35 min.

Mutant D154L showed no dehydrogenase activity under the standard assay conditions. All substrates (including those that are not oxidised by the WT enzyme) and both co-factors were tested but no activity could be detected. Analysis of cell extract and heat-treated protein samples by SDS PAGE (Fig. 5.04) showed that there was no reduction in the amount of soluble GDH after treatment at 70°C for 35 min, indicating that the lack of activity was not as a result of changes to the solubility of the protein.

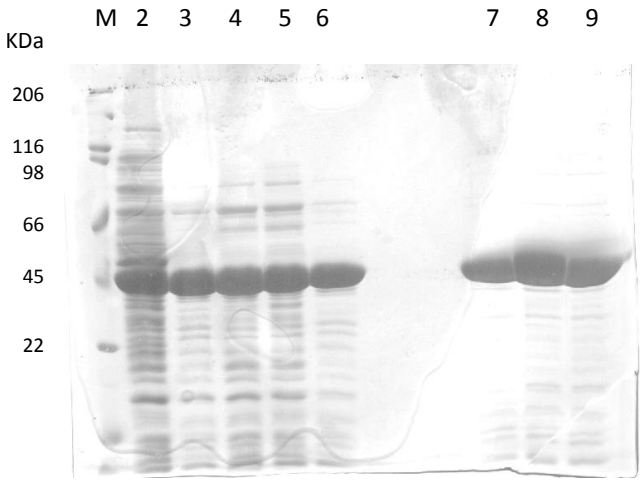


Fig. 5.04. SDS PAGE gel showing soluble expression fractions and heat-treated GDH proteins. M) Broad Range molecular markers (Biorad), Lane 2 shows D154L soluble expression fraction, lanes 3-6 show the soluble fraction subjected to the following heat treatments 3) 60°C/15min, 4) 60°C/30min, 5) 70°C/15min, 6) 70°C/30min. Lanes 7-9 show WT, N89D and Q150E after incubation at 70°C for 35min.

Mutants N89D, N89E and Q150E were all active, but with greatly reduced activity with the standard substrates D-glucose, D-galactose, D-xylose, L-arabinose and D-glucosamine compared to the WT enzyme (Table 5.02). No activity was detected with D-lyxose, L-lyxose, L-xylose, D-arabinose, D-mannose or D-ribose.

	Activity (U/mg)			
Sugar	**WT**	**N89D**	**N89E**	**Q150E**
D-Glucose	55.0	0.9	1.1	1.7
D-Galactose	45.6	0.3	0.2	2.8
D-Xylose	71.0	13.3	4.9	8.6
L-Arabinose	59.6	0.2	0.1	1.3
D-Glucosamine	8.7	0.2	0.1	2.0
D-lyxose	0.0	0.0	0.0	0.0
L-lyxose	0.0	0.0	0.0	0.0
L-xylose	0.0	0.0	0.0	0.0
D-arabinose	0.0	0.0	0.0	0.0
D-mannose	0.0	0.0	0.0	0.0
D-ribose	0.0	0.0	0.0	0.0

Table 5.02. Substrate specificities of the wild type (WT) GDH and mutants N89D, N89E and Q150E. All assays were carried out at as described at 70°C in with 5mM substrate and 0.5mM NADP as a co-factor. One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

Kinetic parameters could only be established for mutants N89D and Q150E with xylose as a substrate due to an observed lag effect in the progress curve of the reaction at low substrate concentrations when the mutants were assayed with D-glucose (Fig. 5.05), D-galactose or D-glucosamine. This effect was present in N89E with all substrates.

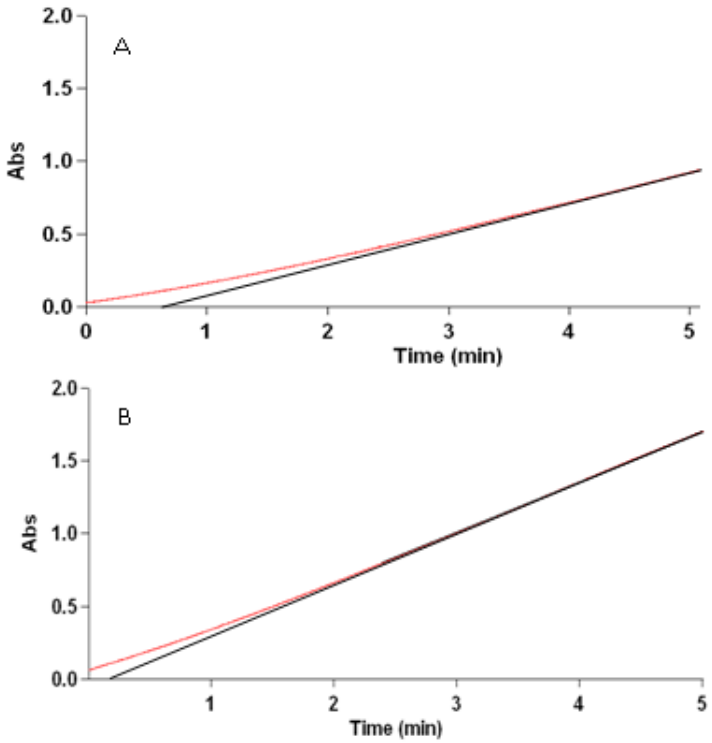


Fig. 5.05. The effect of increasing substrate concentration on the reaction progress of mutant N89D with (A) 5mM glucose and (B) 10mM glucose, preventing the determination of kinetic parameters for most substrates.

5.3.2. Kinetic analysis of WT and mutant enzymes

Kinetic parameters determined for N89D and Q150E showed a dramatic alteration of K_M for D-xylose, which increased from 0.2mM (± 0.004) to 14.5 (± 0.3) and 20.8 (± 0.3) respectively, but with a comparatively small change in V_{max} in both cases (table 5.03).

Enzyme	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
WT	0.2 (± 0.004)	71 (± 0.01)	49	271
N89D	14.5 (± 0.3)	51 (± 0.8)	35	2
Q150E	20.8 (± 0.3)	44 (± 0.6)	30	1

Table 5.03. Kinetic parameters of WT, N89D and Q150E enzymes at 70°C with xylose as a substrate and 0.5mM $NADP^+$ as a co-factor, determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

In contrast to the obvious disruption to xylose binding, determination of kinetic parameters with respect to $NADP^+$ showed that the mutations had not had such a dramatic effect on co-factor binding. K_M was increased slightly from 0.03mM to 0.05mM in the case of Q150E, but there was no observable change with respect to N89D compared to the WT enzyme (Table 5.04). Values of k_{cat}/K_M were affected due to the lowered V_{max}, but were much less significantly reduced compared to N89D and Q150E with xylose, where a 135 and 270 fold decrease was calculated compared to the WT. There was some discrepancy between the V_{max} values obtained for xylose and $NADP^+$ for mutants N89D and Q150E. This observed reduction in V_{max} for $NADP^+$ is most likely due to the enzyme not being saturated with xylose during the assay as the K_M for both mutants is significantly higher than for the WT.

Enzyme	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
WT	0.03 (± 0.001)	54 (± 0.2)	37	1400
N89D	0.03 (± 0.001)	18 (± 0.3)	12	463
Q150E	0.05 (± 0.005)	25 (± 0.8)	17	369

Table 5.04. Kinetic parameters of $NADP^+$ at 70°C with xylose as a substrate at 2mM (WT), 150mM (N89D) and 200mM (Q150E) determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

To investigate these key residues further, 6 more single mutants were constructed and tested with the full range of sugars. These were D154I and D154N, which were apparently soluble but inactive, and N89A, H297F, H297Y and H297W which produced active enzymes.

The design of mutants D154I, D154N and N89A was again to investigate the effect of disrupting hydrogen bonding at the C2 and C3 positions of the natural substrates. Removing the polar residue N89 and replacing it with Ala might remove a hydrogen bond at the C3 position and may allow a C3 epimer to be accommodated. Residue H297, which is donated from the adjacent subunit, has the effect of narrowing one end of the binding cleft. Replacing this residue with the aromatic Phe, Tyr and Trp would presumably alter the amount of available space in the active site at the C6 position of glucose and galactose which would in turn affect which substrates are able to bind.

Mutants N89A and H297F/Y/W all showed activity with the natural substrates; however, xylose was the only sugar that could be used for comparison of kinetic parameters as N89A showed the same lag effect with all other sugars that was observed with the previous mutants (Table 5.05). In contrast to N89D, N89A had a much smaller effect on substrate binding, increasing the K_M for xylose by less than ten-fold (1.7 ± 0.1mM) from the WT value (0.2 ± 0.004mM) compared to the seventy-fold increase observed for N89D (14.5 ± 0.3mM) (Fig. 5.06). Mutants H297F and H297Y had no significant effect on the K_M of xylose, and H297W showed only a moderate increase (Table 5.05).

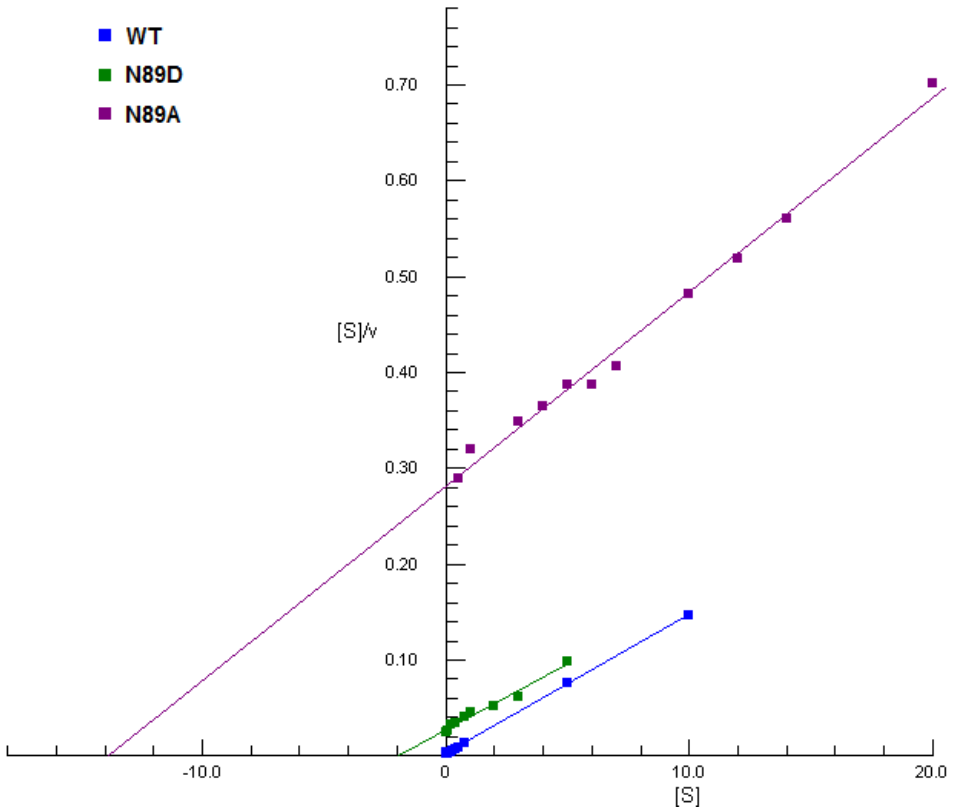


Fig. 5.06. Hanes-Woolf plot for WT GDH and N89D / N89A mutants at 70°C with xylose as a substrate and 0.5mM $NADP^+$ as a co-factor. [S] Is shown in mM and v is U/mg where one unit of enzyme activity is defined by the amount of enzyme required to produce 1µmol of NADPH per min.

Mutation	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
WT	0.2 (± 0.004)	71 (± 0.01)	49	271
N89A	1.7 (± 0.10)	67 (± 2.50)	46	26
H297F	0.2 (± 0.01)	57 (± 0.30)	39	225
H297Y	0.1 (± 0.01)	53 (± 0.60)	37	263
H297W	0.5 (± 0.01)	9 (± 0.20)	6	12

Table 5.05. Kinetic parameters of GDH mutants at 70°C with xylose as a substrate and 0.5mM $NADP^+$ as co-factor, determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

In contrast to previous mutants, the H297 mutants showed normal reaction progress curves with all natural substrates, which allowed the kinetic parameters with D-glucose, D-galactose, D-xylose, and D-glucosamine to be determined in comparison with the WT enzyme (Table 5.06).

With glucose as a substrate, H297F and H297Y both showed a decrease in K_M compared to the WT. K_M remained unchanged for H279W but interestingly, the V_{max} was dramatically reduced compared to WT and other H297 mutants. In contrast, all mutants showed an increased K_M for galactose (the C4 epimer of glucose) compared to the WT. In the case of H297F and H297Y this was coupled with an increase in V_{max} compared to the WT.

The most dramatic decrease in K_M was found with glucosamine, with a ten-fold and eight-fold reduction for H297F and H297Y respectively, in comparison to the WT. Interestingly, mutant H297W showed no activity with glucosamine (2-amino-glucose) or fucose, (6-deoxy-galactose). As with glucose, there was only a slight change in the K_M of fucose when values for the WT and H297F and H297Y were compared (Table 5.06).

Unlike the mutants previously characterised (N89D and Q150E) mutations at H297 appeared to have more of an effect on co-factor binding (Table 5.07). For example, H297F showed a five-fold decrease in K_M for $NADP^+$ with glucose as a substrate compared to the WT and yet remained unchanged when the same parameters were calculated with xylose. In contrast, H297Y showed a two-fold increase in K_M and H297W a slight decrease in K_M when either substrate was used.

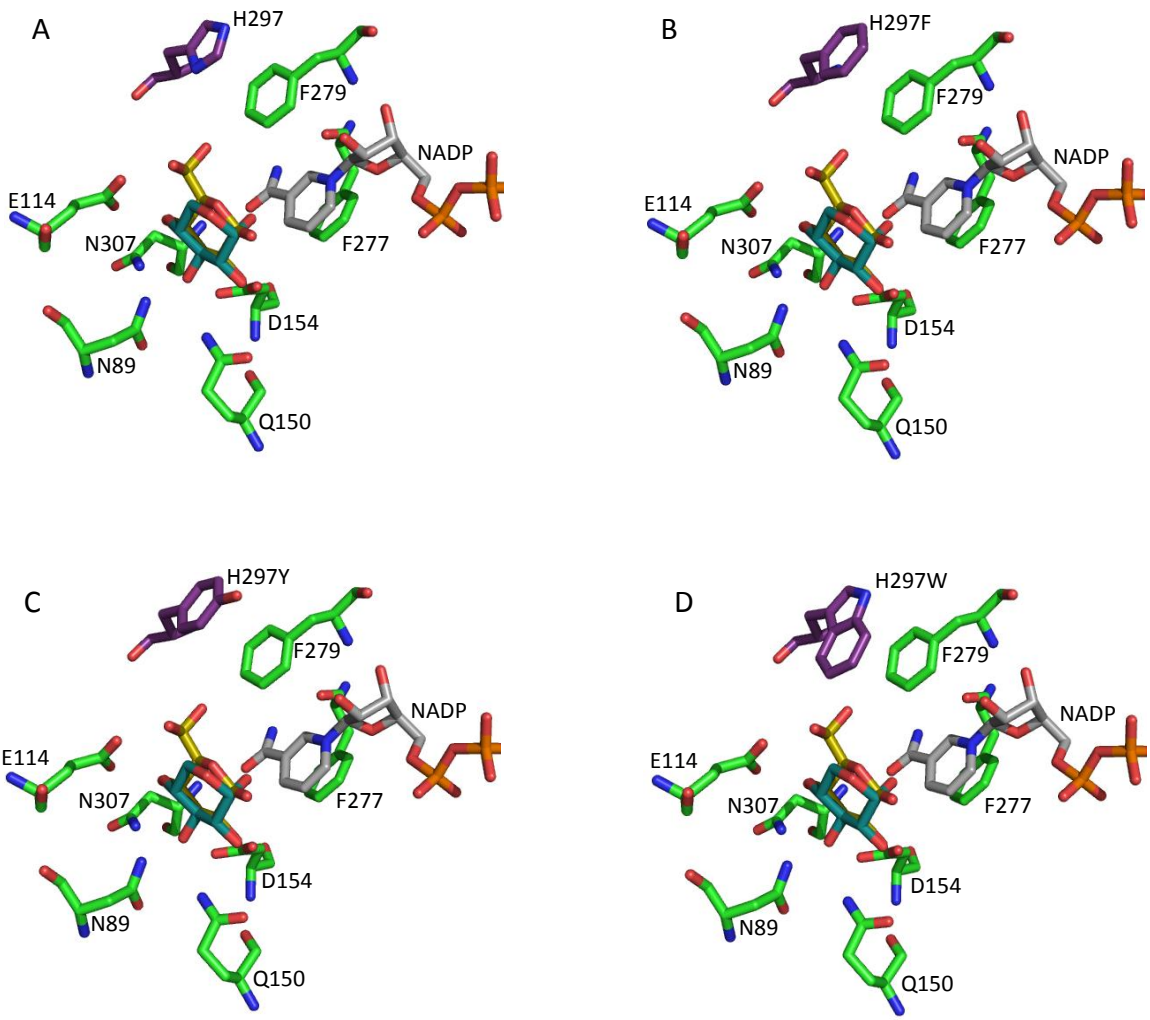


Fig. 5.07. A) Crystal structure of WT glucose dehydrogenase with a molecule of glucose / xylose in the active site. B)/C)/D) GDH structure with probable orientation of residue 297 mutated to phenylalanine (F), tyrosine (Y) and tryptophan (W).

A comparison of models of the GDH active site with residue H297 replaced with Phe, Tyr and Try (Fig 5.07) indicates that the introduction of Try may lead to a narrowing of the binding cleft; possibly explaining the observed reduction in activity with natural substrates (Tables 5.06 and 5.07).

Enzyme	Substrate	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
WT	Glucose	1.5 (± 0.04)	67 (± 0.8)	46	31
H297F	Glucose	0.3 (± 0.01)	46 (± 0.4)	31	102
H297Y	Glucose	0.5 (± 0.01)	61 (± 0.3)	42	86
H297W	Glucose	1.5 (± 0.04)	8 (± 0.1)	6	4
WT	Galactose	0.6 (± 0.01)	59 (± 0.7)	40	64
H297F	Galactose	1.2 (± 0.03)	81 (± 0.5)	56	46
H297Y	Galactose	1.3 (± 0.02)	82 (± 1.2)	56	43
H297W	Galactose	2.6 (± 0.07)	11 (± 0.1)	7	3
WT	Glucosamine	7.2 (± 0.70)	19 (± 1.1)	13	2
H297F	Glucosamine	0.7 (± 0.02)	5 (± 0.04)	3	5
H297Y	Glucosamine	0.9 (± 0.03)	12 (± 0.1)	8	9
H297W	Glucosamine	-	-	-	-
WT	Fucose	0.2 (± 0.01)	61 (± 1)	41	240
H297F	Fucose	0.6 (± 0.02)	90 (± 1)	62	100
H297Y	Fucose	0.2 (± 0.03)	54 (± 1)	37	247
H297W	Fucose	-	-	-	-
WT	Xylose	0.2 (± 0.004)	71 (± 0.01)	49	271
H297F	Xylose	0.2 (± 0.01)	57 (± 0.30)	39	225
H297Y	Xylose	0.1 (± 0.01)	53 (± 0.60)	37	263
H297W	Xylose	0.5 (± 0.01)	9 (± 0.20)	6	12

Table 5.06. Kinetic parameters of WT and H297 mutants at 70°C with a variety of substrates and 0.5mM $NADP^+$ as a co-factor, determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

Enzyme	Substrate	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
WT	Glucose	0.05 (± 0.00)	130 (± 1.10)	89	1931
H297F	Glucose	0.01 (± 0.02)	45 (± 0.02)	31	2196
H297Y	Glucose	0.10 (± 0.02)	34 (± 1.60)	23	232
H297W	Glucose	0.02 (± 0.00)	6 (± 0.40)	5	239
WT	Xylose	0.03 (± 0.003)	54 (± 2.20)	37	1400
H297F	Xylose	0.03 (± 0.002)	56 (± 0.02)	38	1379
H297Y	Xylose	0.07 (± 0.004)	28 (± 0.30)	19	262
H297W	Xylose	0.02 (± 0.000)	7 (± 0.02)	4	225

Table 5.07. Kinetic parameters of $NADP^+$ for WT and H297 mutants at 70°C with glucose at or xylose at approximately 10x K_M, determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974), where one unit (U) of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

In addition to these substrates, activity was also detected with the C5 sugar D-ribose: H297F (12 U/mg), H297Y (0.8 U/mg) and H197W (0.6U/mg) at a substrate concentration of 5mM. No activity could be detected with D-lyxose, L-lyxose, L-xylose, D-arabinose or D-

mannose. K_M and V_{max} were calculated with D-ribose as a substrate for H297F (K_M = 28 ±1.4mM and V_{max} = 24 ±0.6U/mg) but activity was too low to obtain these values for H297Y or H297W.

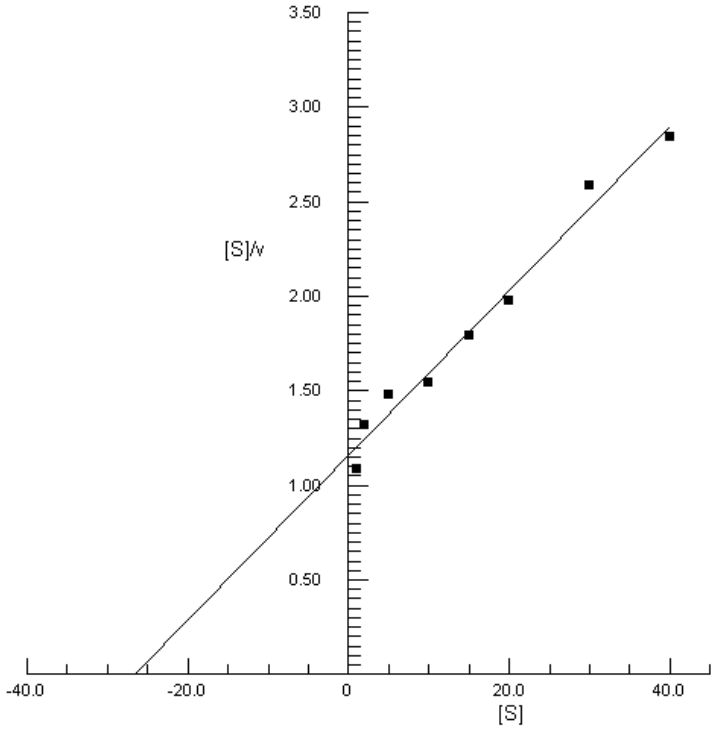


Fig. 5.08. Hanes - Woolf plot for mutant H297F at 70°C under standard assay conditions with ribose as a substrate and 0.5mM $NADP^+$ as a co-factor. [S] is shown in mM and v is U/mg where one unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of NADPH per minute.

5.3.3. Thermoactivity and thermostability

Introduction of mutations that increase the promiscuity of enzymes has been associated with a lowering of thermostability. H297F was chosen for further study as it was the only mutant that showed a significant activity with a non-natural substrate. This mutation also had virtually no effect on K_M values for xylose and $NADP^+$, compared to the WT enzyme, ensuring both enzymes could be assayed under identical conditions. Also, as K_M has been reported to increase with temperature (Scopes, 1995) the substrate with the lowest K_M was selected to ensure that both enzymes would remain saturated at higher temperatures (even if K_M was to increase ten-fold the substrate concentration at 20mM would be ~10x K_M leaving the enzyme 90% saturated).

Thermoactivity assays were performed as described in general Materials and Methods and the effect of temperature on enzyme velocity can be seen in Fig. 5.09. Overall H297F is less

active than the WT; however, the highest rate of reaction for both enzymes was recorded at 92°C.

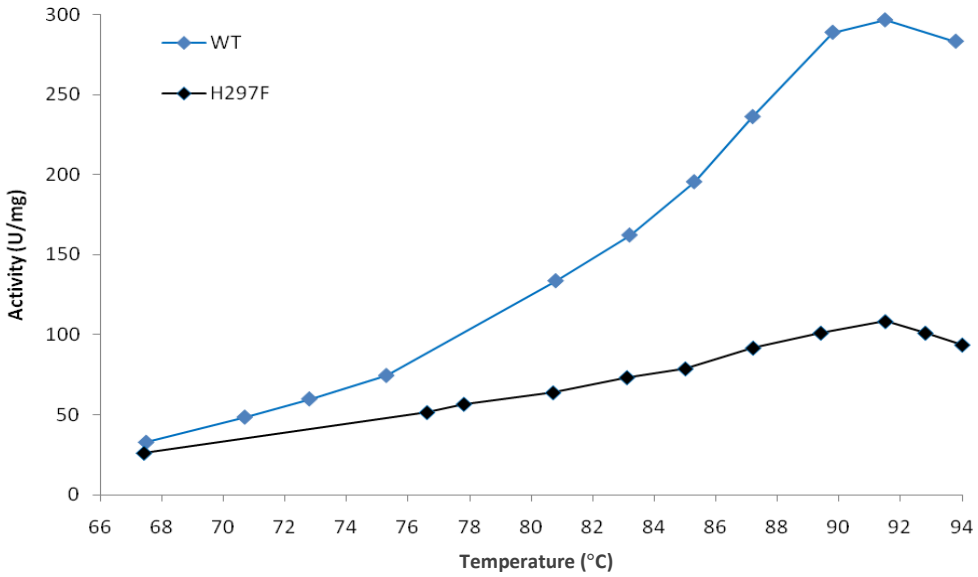


Fig. 5.09. Effect of temperature on enzyme velocity of the WT and H297F enzymes under standard assay conditions with xylose (20mM) and $NADP^+$ (0.5mM).

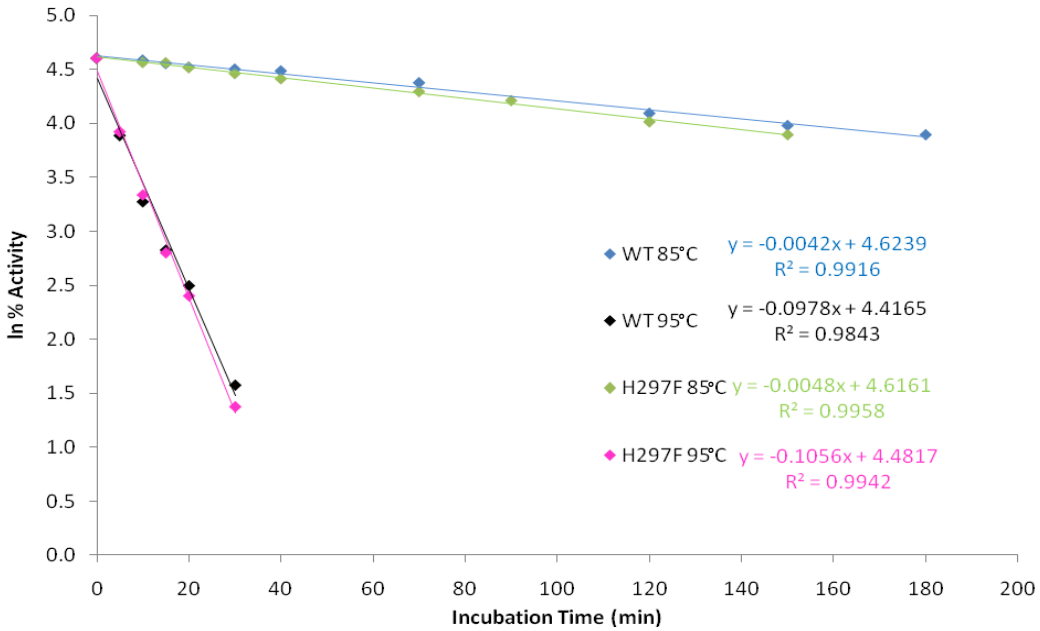


Fig. 5.10. Thermostability of WT and H297F enzymes at 85°C and 95°C,

The thermostability of both enzymes was measured at 85°C and 95°C (Fig. 5.10). The t½ was found to be 150min at 85°C and 5min at 95°C for both enzymes, which indicates that there is no significant difference in the thermostability of H297F and the WT at these temperatures.

5.4. Discussion

The dramatic increase in K_M with respect to xylose for mutants N89D and Q150E, coupled with the lag effect observed with other sugars, confirms that residues N89 and Q150 are critical for hydrogen bonding at the C2 and C3 positions of natural substrates. Mutant N89A resulted in a ten-fold increase in K_M but virtually no change in V_{max}, due to the fact that this replacement would have removed a hydrogen bond at the C3 position, reducing the strength of substrate binding but not effecting the overall rate of catalysis.

Position D154 also appears to be essential, as mutation of this residue to leucine resulted in complete loss of activity despite the enzyme remaining soluble when incubated at 70°C. It is possible that this mutation may have inadvertently caused a slight stabilising effect on the enzyme at the cost of activity, as reported by Bloom *et al* (2006). While these data verify the importance of N89, Q150 and D154 in substrate binding, the mutations made at the positions did not broaden the substrate specificity of the enzyme as predicted.

The introduction of point mutations has in the past been reported to induce changes in promiscuity with little effect on the enzyme's activity with natural substrates (Aharoni *et al*, 2005). The most successful mutants by this standard were those at residue H297, as they conferred activity with ribose, which was not present in the WT enzyme, without causing a detrimental effect on the oxidation of natural substrates. Ribose activity was present at varying degrees depending on the size of the group introduced; the most significant activity was with H297F, but this was reduced by over ten-fold when tyrosine and tryptophan residues were introduced. Restriction of the substrate binding cleft created by histidine at position 297 may have been relieved by the introduction of the unreactive phenylalanine residue. This may have allowed more flexibility in the active site which in turn allowed the ribose and $NADP^+$ to move more freely in the active site, permitting the ring of $NADP^+$ to get close enough to the C1 of ribose to allow hydrogen transfer. In the case of H297Y and H297W the binding cleft may have once again become restricted, resulting in a reduced ribose activity.

Removal of the C6 hydrogen bond for glucose and galactose by introducing Phe or Tyr had an interesting effect on the binding of these two substrates causing the observed decrease (glucose) and increase (galactose) in K_M compared to the WT. In the active site glucose and galactose both interact with E114 but glucose has an additional interaction with N307

because of the alternative C4 positioning. The loss of the additional bond at the C6 position may have resulted in galactose not being bound as tightly as in the WT enzyme, thereby allowing a closer proximity to the $NADP^+$ facilitating hydride transfer and resulting in a slightly increased reaction rate.

Although there are successful examples of rational design studies altering the substrate or cofactor specificity of an enzyme (e.g. Pire *et al*, 2009), the results reported here illustrate the difficulties of this method when mutations are designed around the crystal structure of an enzyme. Enzymes are dynamic and consequently the orientation that mutated active site residues will take in relation to substrates / other residues cannot always be accurately predicted. Without almost saturating a single residue with every possible alternative, the chance of success of finding a mutation with the desired effect through rational design is highly variable. It is, however, important to note that with other methods, such as directed evolution which seems to present a greater success rate, these random mutations only actually result in a 0.5-0.01% chance of conferring beneficial effects; the equivalent of five enzymes with desirable changes in activity in every thousand constructed (Bloom & Arnold, 2009).

It has been reported that activity with a new substrates is often coupled with a decrease in protein stability (e.g. Bloom *et al* 2006, Bloom & Arnold 2009). The thermoactivity of H297F was compared to that of the WT enzyme with an excess of both substrates. Although H297F had a much lower activity overall compared to the WT, the maximum velocity for both enzymes was measured at 92°C. Both enzymes also behaved almost identically when thermostability was measured at 85°C and 95°C. These data demonstrate that although the introduction of single mutations often causes a decrease in enzyme stability it is possible to enhance the promiscuity of an enzyme without affecting its thermostability.

Chapter 6: Investigation of temperature related activity and stability through mutagenesis studies of malate synthase from *S. solfataricus*

6.1. Introduction

One of the most fundamental concepts of enzymology is that enzyme activity varies with temperature. Experimental enzyme characterisation inevitably includes a temperature / activity profile, from which the enzyme is ascribed an 'optimum temperature' (T_{opt}) where the highest velocity for the given reaction is recorded. This is based on a two state or 'classical model' (Fig. 6.01) of enzyme activation and inactivation, in which the enzyme can exist in an active (E_{act}) or irreversibly thermally denatured (X) form. This model incorporates the Arrhenius increase in reaction rate with increase in temperature while at the same time there is a corresponding increase in the rate of thermal denaturation. The associated changes of free energy are described by G^*_{cat} and G^*_{inact} respectively (Daniel *et al*, 2001).

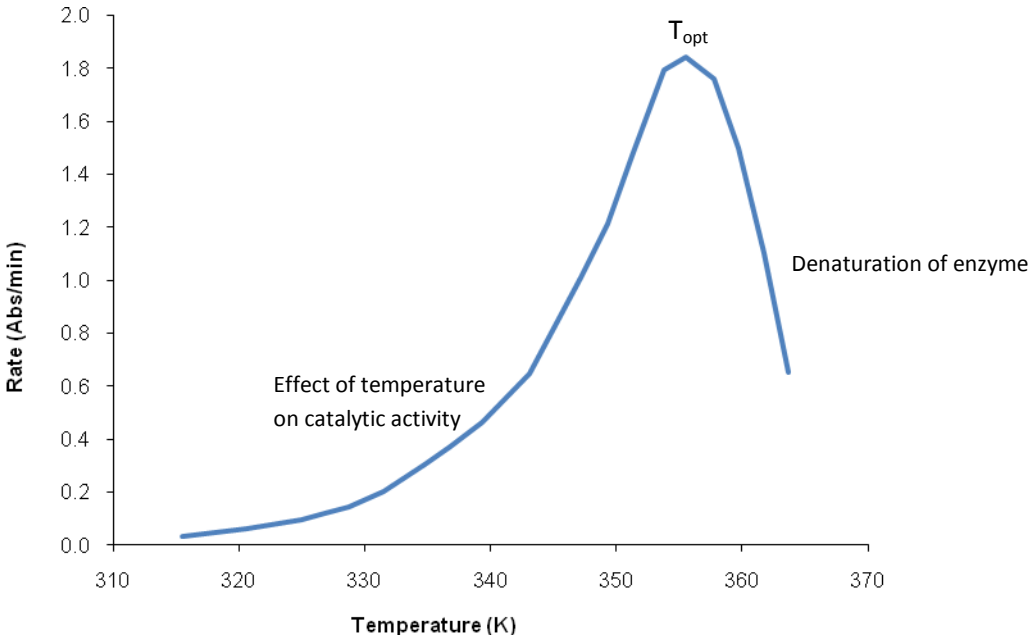


Fig. 6.01. The classical model describing the effect of increasing temperature on enzyme activity. Catalytic rate increases with temperature until a point is reached where the effects of temperature on enzyme stability cause a decrease in catalytic rate. The duration of the enzyme assay also affects the observed catalytic rate and consequently the observed temperature optimum.

There are now many examples of enzymes that have been manipulated for use as biocatalysts or for biotransformation, the objective being to increase or decrease this

'optimum temperature' by altering stability (Serrano *et al*, 1993; Eijsink *et al*, 2004, Eijsink *et al*, 2005; Bloom *et al*, 2006) so that the enzyme can be utilised at a desired temperature where it would not normally operate at maximum velocity.

However, 'optimum temperature' is not an intrinsic enzyme property as it is derived from a complex mixture of both activity and thermal stability effects (Daniel *et al* 2001). The descending limb of Fig. 6.01 arises predominantly from denaturation of the enzyme, which is both time and temperature dependent. Denaturation is relatively slow and therefore at short assay durations (where the effect of denaturation is negligible) the classical model displays no 'optimum temperature'. The result of this is that shorter enzyme assays will result in a higher 'optimum temperature' as shown in the activity vs temperature vs time plot of Fig. 6.02A. This can be described by equation 1;

(Eq. 1)

$$V_{\max} = k_{cat} \cdot [E_0] \cdot e^{-k_{inact} \cdot t}$$

Where V_{max} = maximum velocity of the enzyme, k_{cat} = the catalytic rate constant of the enzyme, $[E_0]$ = total concentration of enzyme, k_{inact} = thermal inactivation rate constant and t = assay duration. The variation of the rate constants k_{cat} and k_{inact} with temperature are defined by;

$$k_{cat} = \frac{K_B T}{h} \cdot e^{-\left(\frac{\Delta G^*_{cat}}{RT}\right)}$$

and

$$k_{inact} = \frac{K_B T}{h} \cdot e^{-\left(\frac{\Delta G^*_{inact}}{RT}\right)}$$

Where k_B = Boltzmann's constant, R = gas constant and h = Planck's constant.

Experimental observations show that the temperature dependent behaviour of enzymes cannot be fully described in terms of these two parameters, due to the much greater reduction in activity with temperature than can be accounted for by denaturation (Thomas & Scopes, 1998). Findings such as these have lead to the formulation of a new model of

temperature dependent behaviour, the Equilibrium Model (Daniel *et al,* 2001). In this new model, the enzyme rapidly converts between active (E_{act}) and inactive (E_{inact}) forms, the latter of which can undergo irreversible denaturation (X).

$$E_{act} \overset{K_{eq}}{\rightleftharpoons} E_{inact} \xrightarrow{k_{inact}} X$$

In this situation, the concentration of the active enzyme at any time point can be defined by equation 2:

(Eq. 2)

$$(E_{act}) = \frac{(E_0) - (X)}{1 + K_{eq}}$$

where K_{eq} is the equilibrium constant between active and inactive forms of the enzyme (K_{eq} = (E_{inact})/(E_{act})). K_{eq} is now a further temperature-dependent property of an enzyme in addition to k_{cat} and $k_{inact.}$ The variation of K_{eq} with temperature can be described by equation 3

(Eq.3)

$$\ln(K_{eq}) = \frac{\Delta H_{eq}}{R}\left(\frac{1}{T_{eq}} - \frac{1}{T}\right)$$

where ΔH_{eq} is the enthalpic change associated with the conversion of an active to an inactive enzyme, and T_{eq} is the temperature at the midpoint of transition between the two forms. At T_{eq}, [E_{act}] = [E_{inact}], therefore K_{eq} = 1, ΔG^*_{eq} = 0 and $\Delta H_{eq} = T_{eq}\,\Delta S_{eq}$ (where ΔS_{eq} is the change in entropy associated with the equilibrium) (Daniel *et al* 2008). By incorporating these new parameters, a mathematical formulation has been developed where the overall effect of temperature on enzyme velocity can be described by equation 4.

(Eq. 4)

$$V_{\max} = \frac{k_B T\, e^{-\left(\frac{\Delta G^{*}_{cat}}{RT}\right)} E_0\, e^{\left(\frac{-k_B T\, e^{-\left(\frac{\Delta G^{*}_{inact}}{RT}\right)} e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{Teq}-\frac{1}{T}\right)}{R}\right)} t}{h\left(1+e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{Teq}-\frac{1}{T}\right)}{R}\right)}\right)}\right)}}{h\left(1+e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{T_{eq}}-\frac{1}{T}\right)}{R}\right)}\right)}$$

The effect of incorporating the equilibrium between E_{act} and E_{inact} can clearly be seen in the profile of the activity vs temperature vs time plot shown in Fig. 6.02B, where there is now a temperature optimum at time zero which is not seen in the classical model.

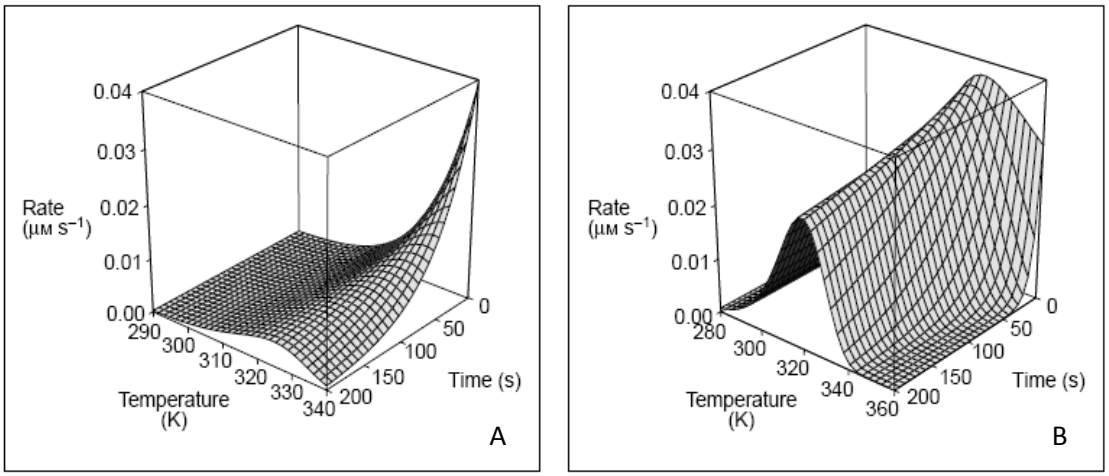


Fig. 6.02. (Taken from Daniel *et al*, 2001). A: Temperature / activity / time plot illustrating the classical model. Denaturation is relatively slow, therefore at short assay durations (where the effect of denaturation is negligible) the classical model displays no 'optimum temperature'. The result of this is that shorter enzyme assays will result in a higher 'optimum temperature'. B: Temperature / activity / time plot illustrating the Equilibrium Model. Rapid conversion between the active and inactive forms of the enzyme results in an initial rate temperature optimum.

The effect of temperature on enzyme activity over time has been recorded for over 30 enzymes and the data generated fitted the Equilibrium Model allowing values for all the parameters T_{eq}, ΔH_{eq}, ΔG^{*}_{cat} and ΔG^{*}_{inact} to be determined (Peterson *et al* 2004, Lee *et al* 2007, Daniel *et al* 2010). Experimental observations show that the equilibrium between the active and inactive forms of the enzyme is rapid, as at high temperatures the enzyme displays a decreased activity at time zero in the assay before thermal denaturation could have occurred (Peterson *et al* 2007, Daniel *et al* 2008), indicating that k_{eq} and k_{inact} are independent of each other. There is evidence to suggest that the E_{inact} form of the enzyme is likely to be a result of reversible conformation changes in the active site rather than significant unfolding of the protein, (Peterson *et al* 2004, Peterson *et al* 2007, Daniel *et al* 2009, Daniel *et al*, 2010). Although the exact nature of these changes is unknown (Daniel *et al* 2009) it raises the idea that mutations could be introduced to manipulate the position of the $E_{act} \rightleftharpoons E_{inact}$ equilibrium without affecting the overall thermal stability of the enzyme. This has major implications for enzyme engineering for biotechnology, as making an enzyme more thermostable (increasing ΔG^{*}_{inact}) will not guarantee an increase in thermoactivity if T_{eq} remains unaffected (Peterson *et al* 2007, Daniel *et al* 2009).

The enzyme chosen for this study is a recombinant malate synthase from the thermophilic archaea *Sulfolobus solfataricus.* Malate synthase catalyses the essentially irreversible condensation of glyoxylate and acetyl-CoA and has previously been cloned and expressed in *E. coli* and found to be a thermostable homodimer of 189KDa, with K_M values for glyoxylate and acetyl-CoA both in the micromolar range (60 (±2) µM and 2 (±0.09) µM respectively) (Chapter 4). During initial characterisation of this enzyme with respect to temperature, the maximum velocity of the enzyme was found to occur at approximately 80°C. There was, however, a loss of over 30% of activity when the temperature was increased to 85°C. In the classical or 'two state' model loss of activity with temperature would be accounted for by irreversible thermal denaturation. However, as the activity at each temperature was calculated from the first 20 seconds of the enzyme reaction, this dramatic reduction in activity over the 5°C temperature change cannot be accounted for by irreversible enzyme denaturation, which occurs over a period of several minutes rather than seconds.

The aims of this study are to characterise the temperature dependent behaviour of the wild type malate synthase based on the assumptions of the Equilibrium Model. To further

this, mutations will be introduced to specific, non-substrate binding residues in the enzyme to investigate whether it is indeed possible to affect T_{eq} and ΔG^{*}_{inact} independently.

6.2. Materials and methods

6.2.1. Malate synthase assay

Assays were carried out as described previously on a Varian spectrophotometer, with a single Peltier-effect cell, using a quartz cuvette of 10mm path length.

6.2.2. Determination of apparent optimum temperature

The standard assay as described in general Materials and Methods was performed at a range of temperatures between 50-90°C. The cuvette containing the buffer and a mixing wand was equilibrated to the desired temperature before the addition of the substrates. These were again allowed to reach the desired temperature (measured with a Whatman DT100 thermocoupler) before the addition of 2µl enzyme. This was thoroughly mixed for no more than 2 seconds before recording was started. Initial rates were obtained by calculating the absorbance change per second and taking an average over the first ~20sec of the reaction.

6.2.3. Calculation of Equilibrium Model parameters

Data processing

The processing of enzyme assay data was performed as described by Peterson *et al* (2007) using the MATLAB-based Equilibrium Model data processing application on a Compact Disc (CD-ROM). Absorbance data from the Varian kinetics software package were converted into progress curves of product concentration (M) versus time (s) in Microsoft Excel, producing a single worksheet containing triplicate data for each temperature. These data were loaded into the Equilibrium Model application, which employs a fitting algorithm based on a least squares approximation run over multiple iterations to identify a set of thermodynamic parameters that best describe the experimental data within the confines of the Equilibrium Model (Daniel *et al* 2009). The standard starting values used for the first iteration are ΔG^*_{Inact} (95 kJ mol^{-1}), ΔG^*_{cat} (80 kJ mol^{-1}), ΔH_{eq} (120 kJ mol^{-1}), T_{eq} (320 K).

The resulting parameters were used to generate a simulated 3D plot of increase in product concentration versus temperature versus time, which was then compared to the corresponding 3D plot generated from smoothed raw assay data to ensure the validity of the final Equilibrium Model parameters. The figures in brackets in the tables are the standard deviations of the fit of the data to the model (Peterson *et al* 2007).

6.2.3. Modelling of malate synthase structure

Modelling of the malate synthase structure was performed using Modeller9v5. A suitable template protein was determined by searching the PDB with the sequence of *Sulfolobus solfataricus,* producing a list of several potential template sequences. The sequence with the highest identity (PDB: 1d8c) was used to generate five models of malate synthase. Conserved active site residues based on alignment with this template sequence were discounted as mutagenesis targets.

6.2.5. Selection of mutants

Potential mutagenesis targets were selected from an alignment of archaeal malate synthase sequences. These sequences were ordered in terms of optimum growth temperature and residues selected when there appeared to be a conservative pattern of change with increasing growth temperature as shown in Fig. 6.03 (see appendix 5 for full alignment).

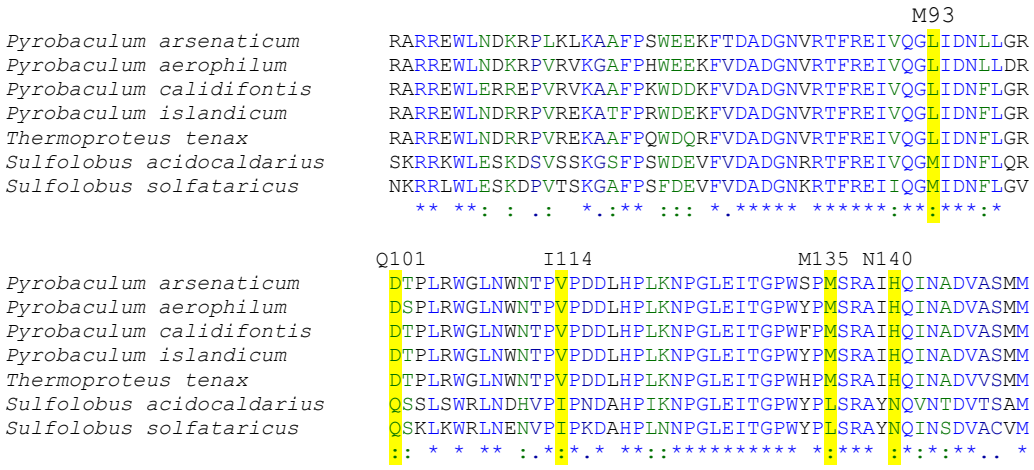


Fig. 6.03. Alignment of archaeal malate synthase sequences based on optimum growth temperature, these are; *Pyrobaculum arsenaticum* (100°C), *Pyrobaculum aerophilum* (95°C), *Pyrobaculum calidifontis* (95°C), *Pyrobaculum islandicum* (90°C), *Thermoproteus tenax* (90°C), *Sulfolobus acidocaldarius* (80°C), *Sulfolobus solfataricus* (80°C). Residues were selected when there appeared to be a conservative pattern of change with increasing growth temperature.

6.2.6. Mutagenesis of malate synthase

The wild type (WT) malate synthase gene had been previously cloned into pET19b as described in chapter 3. Primers designed for mutagenesis are shown in Table 6.01 and the reactions were carried out using Phusion™ polymerase (Finnzymes) and the thermal cycle shown in table 6.02 in an Eppendorf Mastercycler gradient PCR machine.

Mutant	Codon change	Forward primer
M93L	atg →ctc	ggaaatcattcaagggctcatagataacttcctcgg
N140H	aat→cat	ctgagtagggcgtatcatcaaataaactcagacg
I769L	ata→ctc	cccaagtacctcatgaatagcgggaag
L776I	ctc→ata	gcgggaagataatattcatactctctcc
M93A	atg→gct	ggaaatcattcaagggctatagataacttcctcgg
Q101A	cag→gtc	cctcggggtagtctcaaagctaaaatggagg
I114A	ata→gct	gaaaatgtaccagctccaaaggacgcacatcc
L135A	ctg→gct	ggaccttggtatcctgctagtagggcgtataatc
N140A	aat→gct	gagtagggcgtatgctcaaataaactcagacg
L291A	ctt→gct	ggacttaagataggtacagctaagctcgctctcc
P469A	ccc→gct	gaggcagggagcaaagcttaagggcaag

Table 6.01. Forward primers designed for mutagenesis of the *S. solfataricus* WT malate synthase in pET 19b. Reverse primers are direct reverse complement sequences and so are not shown here.

Temperature	Time (min)	Cycles
98°C	2.0	1
98°C	0.5	12
55°C – 65°C	1.0	
72°C	6.0	
72°C	4.0	1
4°C	HOLD	-

Table 6.02. Thermal cycle used for site-directed mutagenesis reaction. Annealing temperature was selected based on the Tm for each pair of primers using the following formula for Tm calculation; Tm = 81.5 + 0.41 (%GC) – 675/N - % Mismatch (N = Primer length in bases)

PCR products were digested with Dpn1 (Promega) for 2-3h and then the remaining plasmid DNA purified with Wizard PCR Cleanup Kit (Promega). 15µl of each reaction was sent for sequencing (Geneservice, Cambridge, UK) using T7F / T7R for mutations in the first or last

800bp of the gene or a specific internal primer (see Materials and Methods chapter 4) to confirm the mutagenesis reactions were successful

If sequencing confirmed the desired mutation had been introduced, the plasmid was transformed into Rosetta™ competent cells (Invitrogen) and grown on LB Agar plates containing carbenicillin (50µg/ml) and chloramphenicol (34µg/ml). Cells containing the construct were used to inoculate 750ml of Overnight Express™ medium (Novagen) and grown for 24h at 37°C in a 2L baffled flask to induce protein expression.

6.2.7. Purification of enzymes

Cells were harvested by centrifugation at 5000 x g for 10 min and resuspended at 0.2g/ml of 50mM Tris-HCl, pH8.5. After 10 min at room temperature the cells were lysed using a cell disruptor (One shot model, Constant Systems, Warwick, UK) and the cell debris removed by centrifugation at 12520xg for 30 min. This cell extract was then incubated at 70°C for 10 min to remove the majority of *E. coli* proteins before being subjected to further purification by FPLC. Samples were filtered through a 0.22µm filter (Milipore, Herts, UK) before loading onto the FPLC. Anion exchange chromatography on two 5ml HiTrap Q Sepharose columns (GE Healthcare, Bucks, UK) using 50mM Tris / HCl (pH8.5) and an elution gradient of 0-1M NaCl and a flow rate of 1.5ml / min. Fractions containing the highest activity were then pooled and concentrated by centrifuging samples at 2880 x g for 5 min using an Amicon Ultra-40 centrifugal tube (Milipore). Gel filtration of the pooled fractions was carried out on a Superdex 200 10/300GL (10mm x 300mm) column (GE Healthcare) using 50mM Tris / HCl (pH8.5) with a flow rate of 0.5ml / min.

6.3 Results

6.3.1. Determination of apparent optimum temperature of the WT malate synthase

Enzyme assays were carried out at a range of temperatures as described to ensure data could be collected beyond the apparent optimum temperature. These data are critical for the calculation of T_{eq}, but would not be easily obtainable if the optimum temperature was 90°C or above as the equipment used had an upper temperature limit of ~93°C. Fig. 6.04 shows the initial rate of reaction (taken over 20s) of the WT enzyme with increasing temperature and an excess of both substrates. From this, the maximum velocity of the enzyme was recorded at approximately 81°C, which is within the range required for accurate data collection.

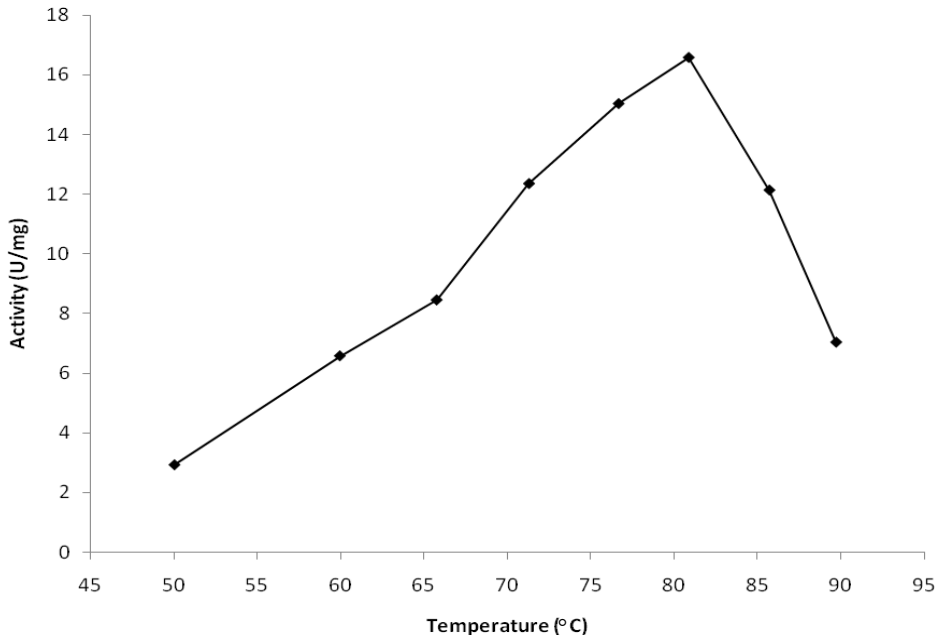


Fig. 6.04. Activity of wild type (WT) malate synthase under standard assay conditions with increasing temperature. One unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

6.3.2. Effect of temperature on K_M of WT malate synthase

K_M has been reported to increase with temperature (Scopes, 1995); therefore kinetic parameters were determined with glyoxylate at 60°C, 70°C and 80°C to ensure that the enzyme would remain saturated at these temperatures. From Fig. 6.05 and Table 6.03, it can be seen that K_M decreases slightly with increasing temperature. While this effect has

not previously been described, for the purpose of this study it implies that the enzyme will remain saturated under the assay conditions at higher temperatures.

Temperature (°C)	K_M (µM)	V_{max} (U/mg)
60	67 (± 2.6)	5.4 (± 0.1)
70	60 (± 3.5)	9.7 (± 0.1)
80	49 (± 0.8)	15.3 (± 0.1)

Table 6.03. Kinetic parameters of WT malate synthase at 60°C, 70°C and 80°C with an excess of acetyl-CoA (0.14mM) or glyoxylate (10mM). Values were determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

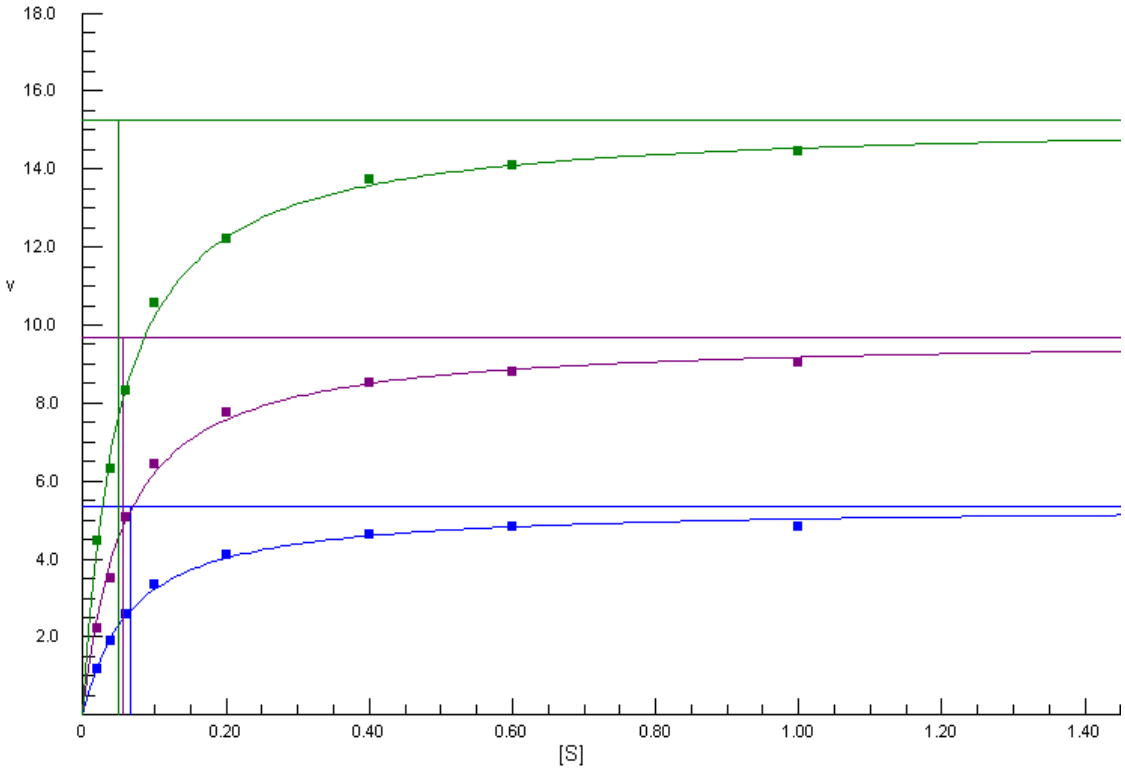


Fig. 6.05. [S] vs Activity plot of WT malate synthase at 60°C, 70°C and 80°C with 0.14mM acetyl-CoA and glyoxylate 0.1 - 1mM. The concentration of substrate [S] is shown in mM and velocity (v) in U/mg where one unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

6.3.3. Calculation of the T_{eq} parameters for the WT malate synthase

Data were collected between 50-90°C and analysed as described in Materials and Methods. The final output values from the MATLAB calculation after four iterations are shown in table 6.04.

ΔG^*_{Inact} (kJ mol^{-1})	100.7 (± 0.1)
ΔG^*_{cat} (kJ mol^{-1})	78.1 (± 0.02)
ΔH_{eq} (kJ mol^{-1})	186.2 (± 1.4)
T_{eq} (K)	353.7 (± 0.1)
R^2	0.992

Table 6.04. Parameters calculated using the equilibrium model. Fitting errors are shown in brackets.

The stability of the enzyme shown in the value for ΔG^*_{Inact}, the catalytic activation energy shown by ΔG^*_{cat} and the enthalpic change associated with the conversion of active to inactive forms are all within the range expected, based on values determined from other thermophilic enzymes (Daniel *et al*, 2010). T_{eq} was calculated as 353.7K or 80.7°C, meaning that at the temperature where maximum velocity of the enzyme was previously recorded only 50% of the enzyme is in the active form.

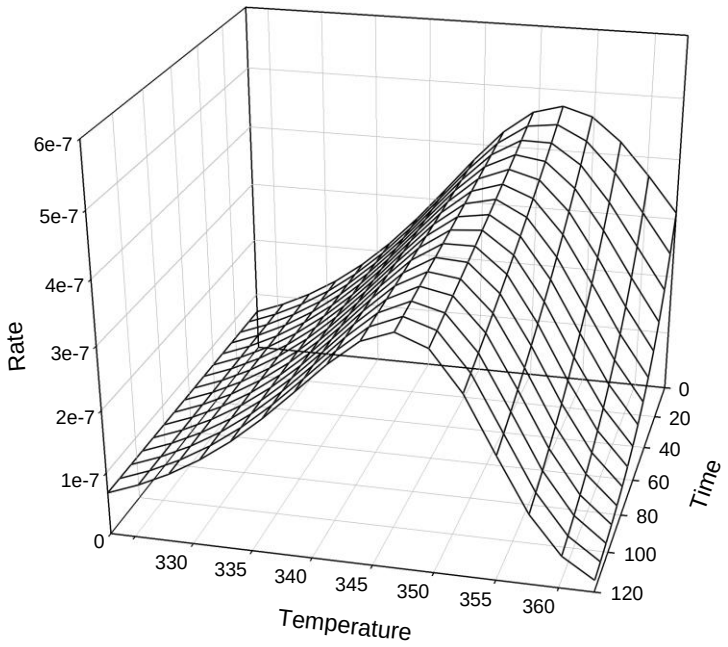


Fig. 6.07. Simulated plot of rate vs time vs temperature using parameters derived from the Equilibrium Model. The graph clearly shows that the enzyme displays an optimum temperature at time zero in accordance with the Equilibrium Model.

The 3D plot of rate, temperature and time in Fig. 6.07 clearly shows that enzyme velocity does not increase exponentially with temperature, instead it exhibits an optimum temperature at time zero, as predicted by the Equilibrium Model.

The fitting errors calculated by MATLAB indicate a normal distribution of data around the calculated values for each of the parameters listed above. The parameters in table 6.04 were used to generate a simulated 3D plot of increase in product concentration versus

temperature versus time, which was then compared to the corresponding 2D plot generated from smoothed raw assay data to ensure the validity of the final Equilibrium Model parameters (Fig. 6.08).

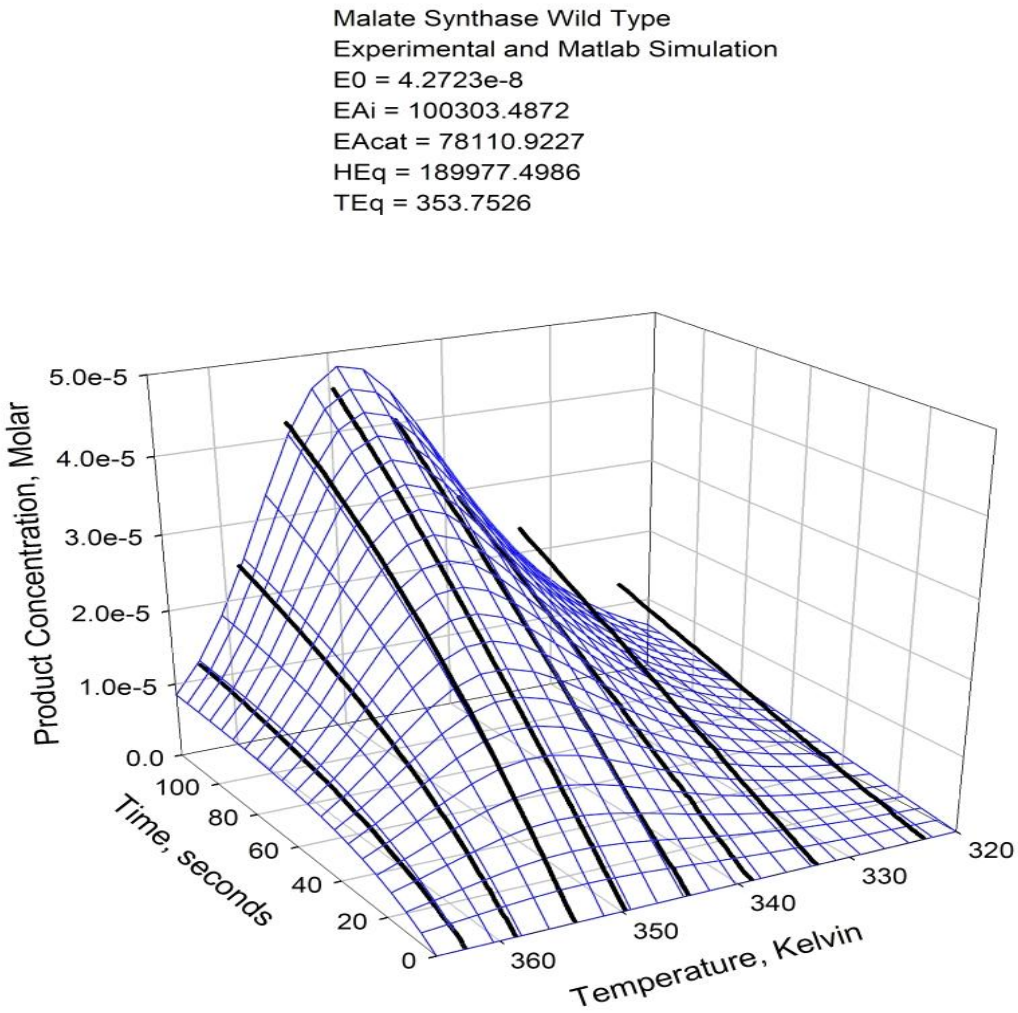


Fig. 6.08. Validation of final Equilibrium Model parameters. Parameters from the WT malate synthase were used to generate a simulated 3D plot of increase in product concentration versus temperature versus time (blue mesh), which was then compared to the corresponding 2D plot generated from smoothed raw assay data (black lines). Parameters are shown as they are inputted into the MATLAB application; EAi = ΔG^{*}_{inact}, EAcat = ΔG^{*}_{cat}, HEq = ΔH_{eq} and TEq = T_{eq}.

Qualitative analysis shows that the raw and simulated data overlay well, (especially at higher temperatures where the effect on T_{eq} is most significant) as suggested by R^2 = 0.992 determined for fitting data to the model and shows that *S. solfataricus* malate synthase is suitable as a model for mutagenesis studies.

6.3.4. Modelling of WT malate synthase structure

A predicted structural model of *S. solfataricus* malate synthase (SsMS) was generated as described in Materials and Methods. As there are no structures solved for archaeal malate synthase enzymes, the sequence selected for a template was that of *E. coli* malate synthase G (*Ec*MSG) which had the highest sequence identity (28%) with SsMS. Interestingly, residues identified in *Ec*MSG as having a function in substrate binding and catalysis appear to be conserved in the thermophilic malate synthase (see also appendix 5).

```
EcMSG     MSQTITQSRLRIDANFKRFVDEEVL------PGT--G---------LDAAAFWRNFDEIVH---D--L
SsMS      VEKLIEELAVEFSDEIRKVINKRRLWLESKDPVTSKGAFPSFDEVFVDADGNKRTFREIIQGMIDNFL

EcMSG     APENRQLLAERDRIQAALDEWHRSNPG-PVK-----DKAAYKSFLRELGYLVPQPERVTVE---TTGI
SsMS      GVQSKLKWRLNENVPIPKDAHPLNNPGLEITGPWYPLSRAYNQINSDVACVMEDEEDASPAWYIPFGS

EcMSG     DSEITSQAGPQLVVPAMNARYALNAANARWGSLYDALYGSD---IIPQE--GAMVSGYD-PQRGEQVI
SsMS      GKTIADVWEGRKNVKLFLSGKAPNPYYEK-GKTYSLNKPRDKWPVIFHRLPGLHLLDFDITLNGKPVP

EcMSG     AWVRRFLDESLPLEN--GSYQDVVAFKVVDKQLRIQLKNGKETTLRTPAQFVGYR-GDAAAPTCILLK
SsMS      AIIVSAVIYTLNNYNSLKSAGSGVYFYLPKTQTPDEALI-IEKILRRIESKLGLKIGTLKLALLYEEV

EcMSG     NNGLHIELQIDA-NGRIGKDDPA---HINDVIVEAAISTILDCEDSVAAVDAE-DKILLYRNLLGLMQ
SsMS      NAGRFFPVILWIFRERLIKSNNGRWDYLGSLIEMWLQEKVLPDPQNITMTSPNMMAYQKYNALMMLLA

                                          338
EcMSG     GTLQRKLNDDRHYTAAD---GSEISLHGRSLLFIRNV-GHLMT------IPVIWDSEGNEIPEGILDG
SsMS      GAKNGEADAAPVGGMAAVMLYPQTDPFGRNRYNLKALRGIKLDKLRERLIGLIFVAEGKVEGKITLE
                                          403
EcMSG     VMTGAIA--LYDLKVQKNSRT-GSVYIVKPKMHGPQEVAFANKLFTRIETML---GMAPNTLKMGIMD
SsMS      IVNGKVKGKLYDMFRQSWVATKEEAYVEAGSKPLRASLDELQKMIDAPINYIEVEGTKLPTVDSGLTP

          427                     453/454/455
EcMSG     EERRTSLNLRSCIAQARNRVAFINTGFLDRTGDEMHSV-MEAGPMLRKNQMKSTPWIKAYERNNVLSG
SsMS      EERALFQKLGLIDERGKITPWVITKEMINTPEKLLFNKELWGGKDLWHS-LYDIP-EGDITPEHVQHA
          505                     531/532/533

EcMSG     LFCGLRGKAQIGKGMWAMP-D--LMADMYSQKGDQLRAGANTAWVPSPTAATLHALHYHQTNVQSVQA
SsMS      FYMAANYGFQLLNGNLAAAIDDYELKQRFMNDLATYRIFTSWLWSVINRDASFTKDGYIKGPKLTKDG

EcMSG     NIAQTE-FNAEFEPLLDDLLTIPVAENANWSAQEIQQELDNNVQGILGYVVRWVEQGIGCSKVPDIHN
SsMS      VIPAEDVLKVTKGTKVKDIFEKLWELHLDW-TYEFYKEQDMRASRRIAETF---GKTNNASTVEEVYK

EcMSG     VALME-DRATLRISSQHIANWLRHGIL--TKEQVQASLENMAKVVDQQNA--------GD---PAYRP
SsMS      VISKAYSSGPFREMSAKEAAQKLAKILNANASEIEEELINLAPRFDRSMAPVIMEILMKEMLYPKYIM

EcMSG     MAGN--FA----NSCAFKAASDLIFLGVKQ-PNGYTEPLLHAWRLREKE----S--
SsMS      NSGKILFILSPLDPERRSKVMDSIFSFRKMIEDKVKRGELDKWVLELYDYVYDNYW
```

Fig. 6.09. Alignment of *S. solfataricus* malate synthase (SsMS) sequence with *E. coli* malate synthase G (EcMSG) (PDB: 1d8c) generated by Modeller9v5. Conserved residues are highlighted in grey, key catalytic residues are either fully conserved (highlighted in black) or semi conserved (highlighted in blue).

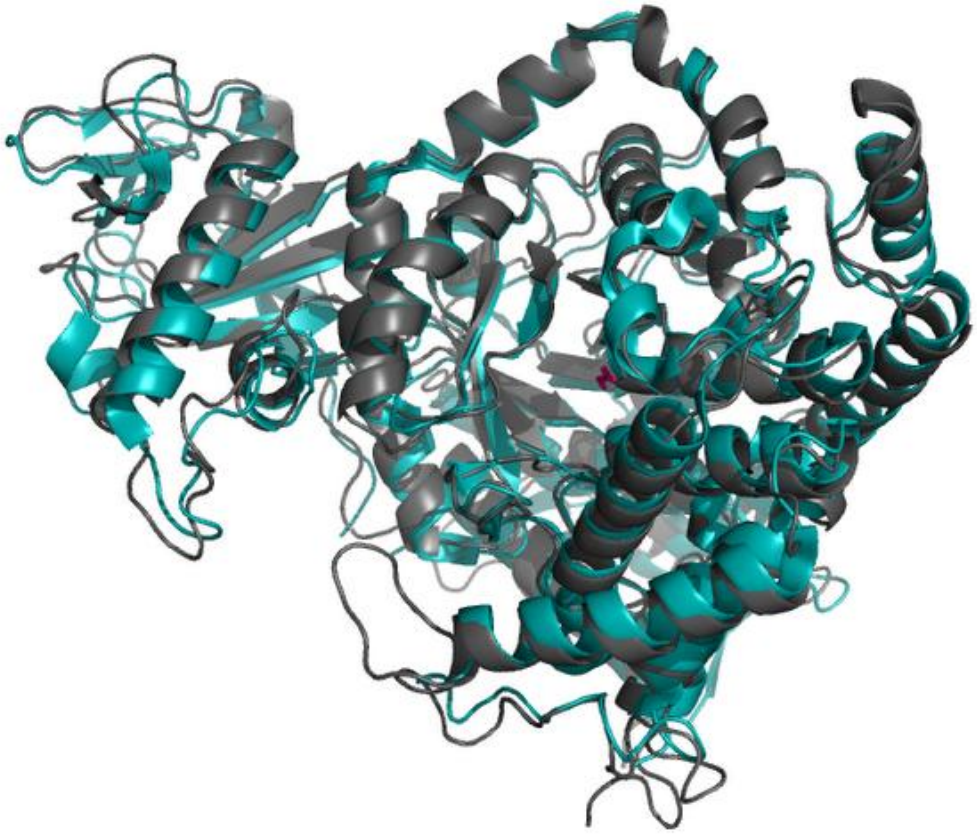


Fig. 6.10. Model of *S. solfataricus* malate synthase (shown in grey) based on the sequence alignment shown in Fig. 6.09 with *E. coli* malate synthase G (shown in teal). A molecule of glyoxylate is shown in the active site in pink

.

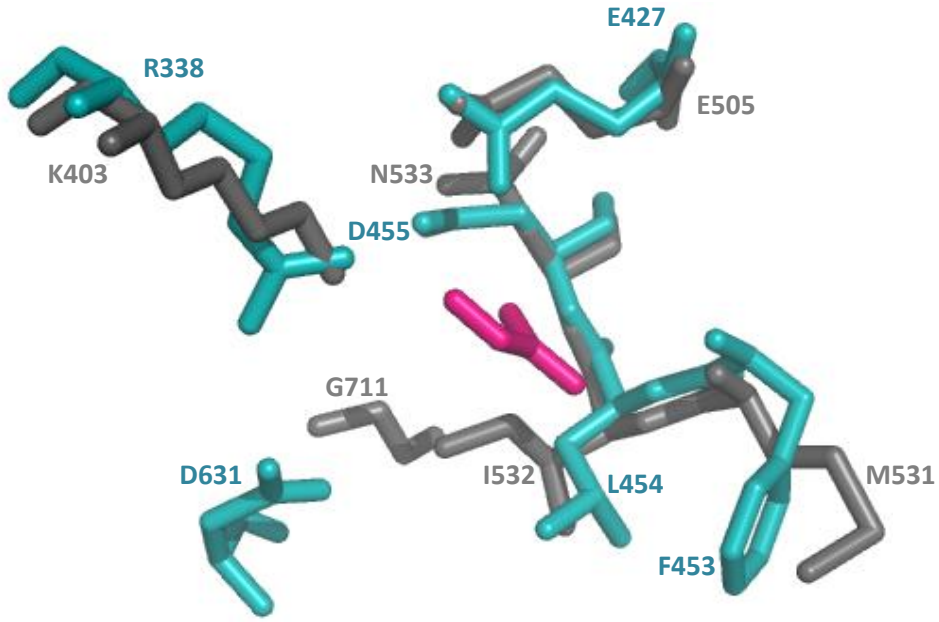


Fig. 6.11. Model of the *S. solfataricus* malate synthase active site (shown in grey) showing predicted key catalytic residues shared with *E. coli* malate synthase G (shown in teal). A molecule of glyoxylate is shown in the active site in pink.

6.3.5. Selection of mutants

From the alignment of archaeal malate synthase enzymes (Appendix 5) several residues were selected as potential targets for mutagenesis as described in Materials and Methods. These residues were mapped onto the model of the WT malate synthase so that their proximity to the active site could be estimated (Fig. 6.12). A selection of residues were then chosen which would provide a range of mutations which were either a relatively short (e.g. P469), medium (e.g. L769/I776) or long distance (e.g. L291) away from the active site so that the effect of both the mutation and its proximity to the active site could be assessed.

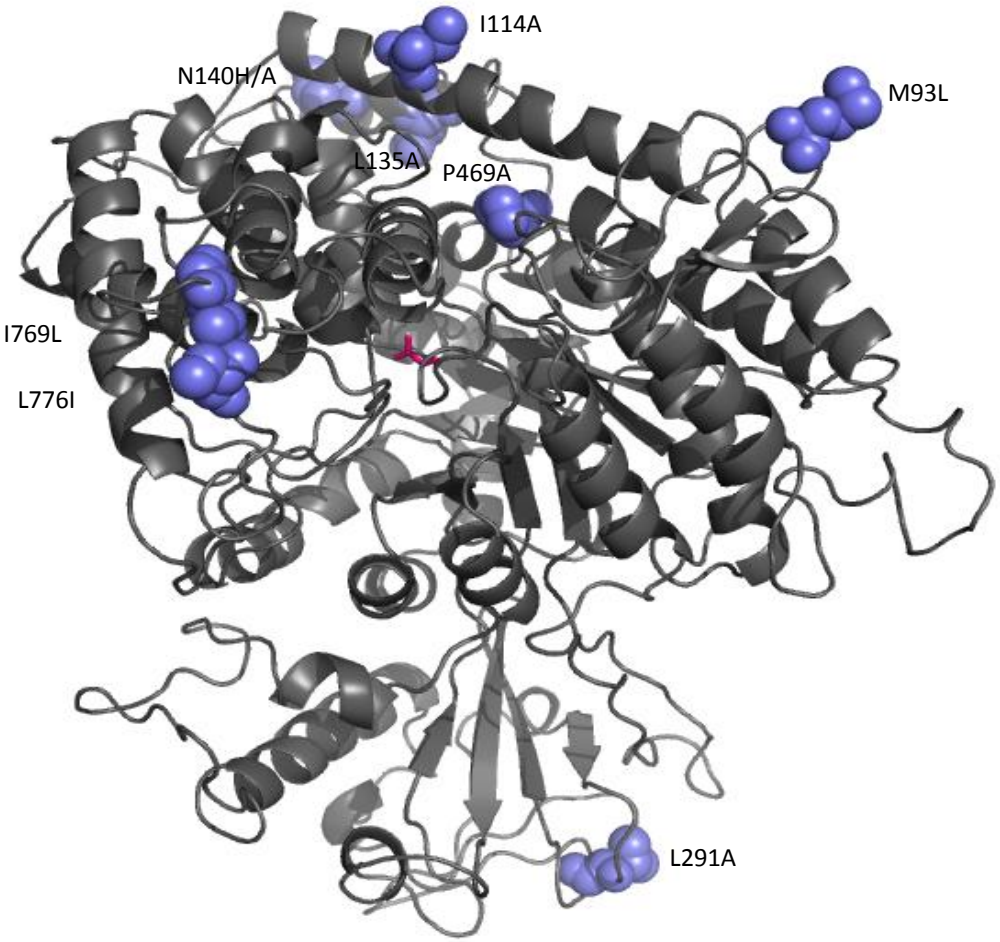


Fig 6.12. Model of *S. solfataricus* malate synthase with the residues selected for mutagenesis highlighted in blue. A molecule of glyoxylate is shown in the active site in pink.

6.3.6. Characterisation of mutant enzymes

Mutant enzymes were constructed and the recombinant proteins expressed as soluble active enzymes and purified as described. Kinetic parameters were determined for all mutants (Table 6.05) before any temperature studies were undertaken to ensure that the mutation introduced did not have a detrimental effect on either K_M (i.e. it became impractical to assay the enzyme under saturating conditions) or V_{max} (the activity was too low to accurately determine any changes in rate associated with temperature).

Calculation of kinetic parameters revealed that all mutants showed an increase in K_M with respect to glyoxylate compared to the WT enzyme. The most significant change was for mutant N140A, where K_M increased from ten-fold for glyoxylate and two-fold for acetyl-CoA. Despite these increases, it is still possible to saturate all enzymes at high temperatures based on the effect on K_M with temperature observed in the WT enzyme (Table 6.05).

	Glyoxylate		Acetyl-CoA	
	K_M (µM)	V_{max} (U/mg)	K_M (µM)	V_{max} (U/mg)
WT	60 (± 2)	12.6 (± 0.17)	1.9 (± 0.10)	13.8 (± 0.05)
M93L	111 (± 3)	16.2 (± 0.07)	3.3 (± 0.03)	17.2 (± 0.04)
I114A	97 (± 4)	13.8 (± 0.10)	1.6 (± 0.06)	14.9 (± 0.03)
L135A	263 (± 6)	18.7 (± 0.15)	2.6 (± 0.23)	18.9 (± 0.17)
N140H	120 (± 5)	16.4 (± 0.12)	4.9 (± 0.24)	19.1 (± 0.33)
N140A	734 (± 27)	22.0 (± 0.14)	4.0 (± 0.16)	22.3 (± 0.12)
I769L	90 (± 3)	13.4 (± 0.05)	3.1 (± 0.12)	14.8 (± 0.16)
L776I	210 (± 4)	7.4 (± 0.05)	2.1 (± 0.12)	7.65 (± 0.03)
I769L / L776I	190 (± 10)	15.0 (± 0.03)	3.3 (± 0.16)	15.6 (± 0.25)
L291A	130 (± 3)	27.5 (± 0.10)	5.8 (± 0.20)	32.1 (± 0.20)
P469A	96 (± 4)	23.4 (± 0.30)	3.5 (± 0.20)	24.1 (± 0.10)

Table 6.04. Kinetic parameters of *S. solfataricus* malate synthase mutants at 70°C with an excess of either acetyl-CoA (0.14mM) or glyoxylate (10mM). Values were determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

The effect of temperature on K_M was also investigated for the double mutant I769L / L776I and the results found to be comparable with those for the WT enzyme, showing a decrease in K_M with increasing temperature (Table 6.06 and Fig. 6.14)

Temperature (°C)	K_M (µM)	V_{max} (U/mg)
60	255 (± 8)	8 (± 0.03)
70	190 (± 8)	15 (± 0.05)
80	139 (± 4)	24 (± 0.03)

Table 6.06. Kinetic parameters of WT malate synthase at 60°C, 70°C and 80°C with an excess of 0.14mM acetyl-CoA and 0.1 - 4mM. Values were determined by the Direct Linear method (Eisenthal & Cornish-Bowden, 1974). One unit (U) of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

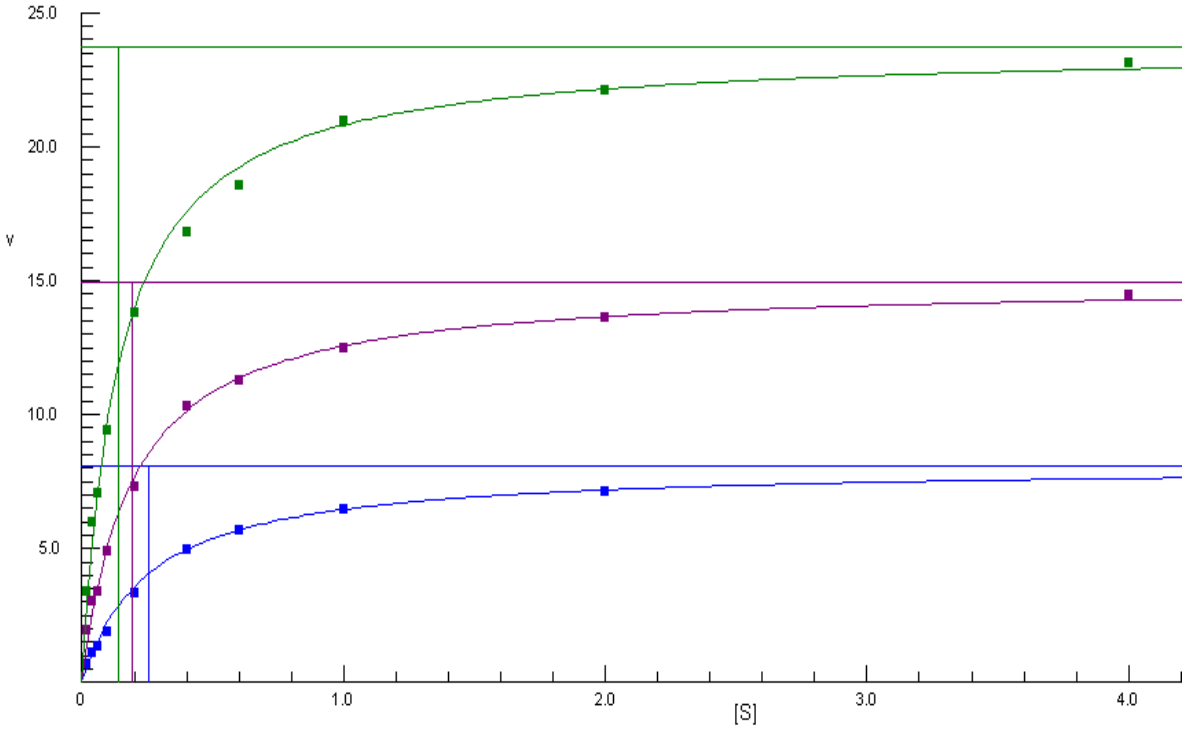


Fig. 6.14. [S] vs Activity plot of mutant I679L / L776I at 60°C, 70°C and 80°C with an excess of acetyl-CoA (0.14mM) or glyoxylate (10mM). The concentration of substrate [S] is shown in mM and velocity (v) in U/mg where one unit of enzyme activity is defined as the amount of enzyme to produce 1µmol of 5-mercapto-2-nitrobenzoic acid per minute.

6.3.7. Calculation of the T_{eq} parameters for malate synthase mutants

Data were collected and analysed as described and the final output values from the MATLAB calculation after four iterations are shown in table 6.07 in comparison with the WT enzyme, ordered by increasing T_{eq} values. The parameters in table 6.07 were used to generate a simulated 3D plot of increase in product concentration versus temperature versus time, which was then compared with the corresponding 2D plot generated from smoothed raw assay data for all mutants to ensure the validity of the final Equilibrium Model parameters (appendix 7). Simulated plots of rate vs time vs temperature using parameters derived from the Equilibrium Model were also created (appendix 6).

	ΔG^*_{Inact} (KJ mol^{-1})	ΔG^*_{cat} (KJ mol^{-1})	ΔH_{eq} (KJ mol^{-1})	T_{eq} (K)	R^2
N140A	101.0 (± 0.6)	74.9 (± 0.4)	153.6 (± 10.4)	348.10 (± 1.9)	0.988
N140H	100.9 (± 0.6)	74.2 (± 0.4)	189.4 (± 12.4)	350.53 (± 1.8)	0.996
L135A	102.1 (± 0.6)	76.6 (± 0.4)	155.8 (± 10.4)	351.21 (± 1.8)	0.995
L291A	102.4 (± 0.6)	71.6 (±0.4)	116.7 (± 7.8)	351.94 (± 1.9)	0.997
M93L	100.7 (± 0.6)	79.0 (± 0.4)	173.9 (± 11.9)	352.46 (± 1.9)	0.993
WT	100.7 (± 0.6)	78.1 (± 0.4)	186.2 (± 12.6)	353.70 (± 1.9)	0.992
I114A	101.6 (± 0.6)	78.9 (± 0.4)	174.3 (± 11.5)	354.04 (± 1.9)	0.996
P469A	102.9 (± 0.6)	69.2 (±0.4)	162.1 (± 10.6)	355.26 (± 1.8)	0.998
I769L	100.5 (± 0.6)	72.4 (± 0.4)	189.3 (± 13.5)	357.53 (± 1.9)	0.995
L776I	101.9 (± 0.6)	70.8 (± 0.4)	160.0 (± 10.2)	358.07 (±1.9)	0.999
I769L / L776I	101.6 (± 0.7)	74.2 (± 0.4)	199.2 (± 14.6)	358.79 (± 2.0)	0.992

Table 6.07. Equilibrium Model parameters. Combined fitting and experimental errors shown in brackets. Experimental error was taken as 0.5% for $\Delta G^*_{inact,}$ ΔG^*_{cat}, and T_{eq} and 6% for ΔH_{eq} as described by Peterson (2007). Data have been ordered in terms of increasing T_{eq} for comparison, the mutants with the greatest change in T_{eq} compared to the WT have been highlighted in blue.

Of the ten mutants analysed, four enzymes N140A, I769L, L776I and I769L/L776I showed a significant change in T_{eq} of ≥4 degrees above or below that of the WT enzyme. All others showed small changes within the range allowed for experimental error and so were not considered to be significant (Fig. 6.16). These enzymes, N140A, I769L and I769L/L776I, showed only slight changes in stability, with the calculated ΔG^*_{Inact} = 101.0 KJ mol^{-1}, 100.5 KJ mol^{-1} and 101.6 KJ mol^{-1} respectively compared to 100.7 KJ mol^{-1} for the WT enzyme. This is not a significant difference given the combined experimental and fitting errors of ~0.6KJ mol^{-1}. L776I showed a greater increase of ΔG^*_{Inact} = 101.9 KJ mol^{-1}, which may be slightly more significant, and so will not be included for further analysis.

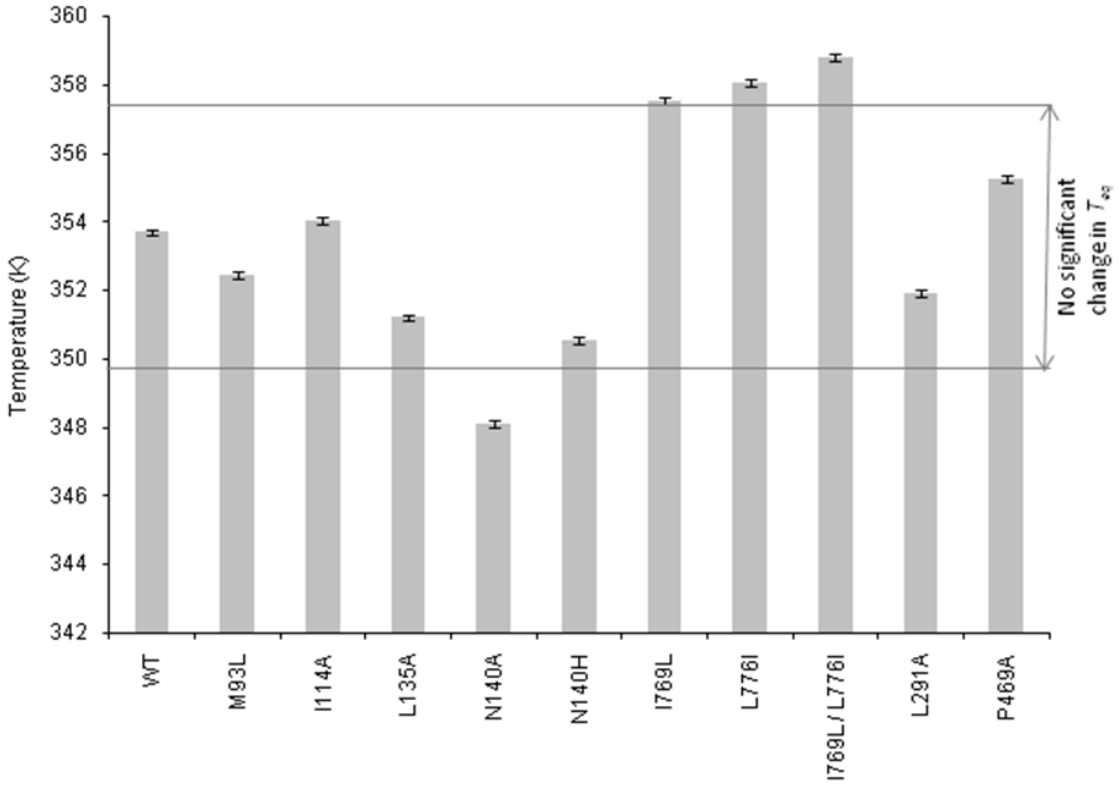


Fig. 6.16. Comparison of calculated T_{eq} values for WT malate synthase and mutants. N140A, I769L, L776I and I769L/L776I showed a significant change in T_{eq} of ≥4°K above or below that of the WT enzyme.

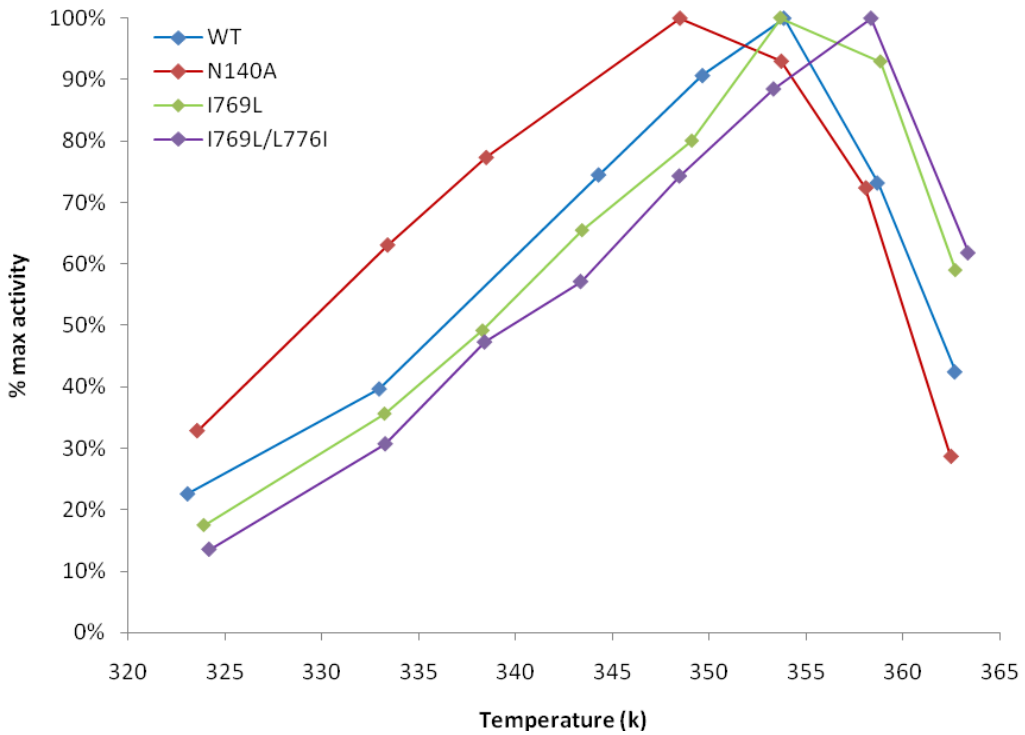


Fig. 6.17. % maximum activity of WT and mutant enzymes with increasing temperature, performed under standard conditions as described. Initial rates (form the first 20 seconds of each reaction) at each temperature were calculated before the values were converted to % maximum for comparison.

The significance of these changes in T_{eq} and ΔG^{*}_{Inact} in terms of the stability and activity of the enzymes can be seen by comparison of % activity of each enzyme between 50-90°C, calculated from the velocity of the reaction over the first 20s of each assay (Fig. 6.17). N140A shows a significant increase in overall activity at lower temperatures (between 50-75°C) compared to the WT and other mutant enzymes and shows maximum activity 5 degrees lower than that of the WT (76°C and 81°C respectively).

There is no change in the temperature at which I769L shows maximum activity compared to the WT, with no significant increase in overall activity at lower temperatures. Above 81°C however, activity is maintained with a loss of just 6% between 81-86°C, compared to the WT where activity rapidly declines with 30% of maximum activity lost over the same 5°C temperature range. The maximum activity of I769L/L776I was found to be 5 degrees higher than WT and 10 degrees higher than N140A, maintaining the highest activity at the maximum temperature of the assay (90°C).

Comparison of activity at ~363K (90°C) (Table 6.08 shows that N140A retains 29% of maximum activity at this temperature compared to 42% for WT enzyme, 59% for I769L and 62% for I769L / L776I.

	% Max activity			
Temperature (k)	**WT**	**I769L**	**I769L/L776I**	**N140A**
324	18%	18%	14%	33%
333	40%	36%	31%	63%
338	55%	49%	47%	77%
344	75%	65%	57%	85%
349	91%	80%	74%	100%
354	100%	100%	88%	93%
358	73%	93%	100%	72%
366	42%	59%	59%	29%

Table 6.08. % Maximum activity of WT malate synthase and four mutants with increasing temperature, calculated from the velocity recorded over the first 20s of the reaction with an excess of both substrates.

6.4. Discussion

Over the past two decades, an increasing amount of research has focused on the alteration of enzyme activity at a specific temperature through mutagenesis. This has been done by a range of methods including site-directed mutagenesis (Eijsink *et al,* 2004), protein design (Serrano *et al*, 1993) and directed evolution (Giver *et al,* 1998; Bloom & Arnold, 2009; Walterink-van Loo *et al*, 2009). In contrast to these studies, the work presented here demonstrates that site-directed mutagenesis can be used to improve both high temperature (I769L and I769L/L776I) and low temperature activity (N140A) without an associated change in thermostability.

All mutations resulted in an increase in K_M, which ranged from moderate in the case of I769L to an extreme in the case of N140A. However, these increases were not always associated with changes in T_{eq} or global stability. For example, mutation M93L caused an increase in the K_M of both glyoxylate and acetyl-CoA, but appeared to be completely neutral with respect to the enzyme's T_{eq}, ΔG^*_{Inact} and ΔG^*_{cat}, as calculated values for all these parameters were almost identical to the WT.

The comparison of parameters from N140A and the WT enzyme provides good evidence that T_{eq} is predominantly affected by small, localised changes (in this case around the active site – given the ten-fold increase in K_M) and not global unfolding caused by denaturation, as there is no significant change in the stability of N140A compared to the WT. Interestingly, mutants P469A and L291A also highlight the independence of T_{eq} and thermostability as they both showed a significant increase in stability (~2kJmol^{-1}) without an accompanying raise in thermoactivity, as T_{eq} remained virtually unchanged compared to the WT enzyme. Mutants I769L and I769L/L776I illustrate that these localised changes can affect T_{eq} without necessarily causing a dramatic increase in K_M.

Residue N140 may be a potential 'hot spot' in relation to T_{eq}, as mutation to His or Ala resulted in a decrease in T_{eq}, although more significantly in the case of N140A. The decrease of ~2.5°C between the two mutants is made more striking by the fact that the activation energy and stability remain completely unchanged but K_M for glyoxylate increases dramatically from two-fold the WT value (N140H) to ten-fold (N140A) indicating that this change must be close to the active site of the enzyme. The effect of introducing

different residues at this position on the E_{act}/E_{inact} equilibrium would therefore be extremely interesting to explore.

Overall, these data support the proposals of the Equilibrium Model that an active and inactive form of an enzyme exist in a rapid reversible equilibrium that is influenced by small, localised changes in structure rather than global unfolding, and that the constant K_{eq} defining this equilibrium is completely independent of k_{inact}, the rate constant that describes the irreversible thermal inactivation of the enzyme to a denatured state.

While significant findings have been made from a relatively small number of mutants, this work does not represent an exhaustive study of mutations of malate synthase. While the structural model produced was useful to ensure that conserved active site residues remained unchanged, but the lack of any detailed structural information has limited the conclusions that can be drawn from these data in relation to the proximity of the T_{eq} modifying residues to the active site. The conserved active site residues between *E. coli* and the thermophilic malate synthase enzymes (with the exception of D631 which is involved in Mg^{2+} binding) suggests that a similar method of substrate binding exists, even though the actual mechanism for catalysis may vary, as the requirement for Mg^{2+} is not shared by the thermophilic enzymes.

There is an ongoing search for enzymes that can be exploited as biocatalyts for a wide variety of applications, including production of fine chemicals and pharmaceuticals. Compared with traditional organic synthesis, biocatalysts are more precise, allowing the production of stereospecific compounds, generally with fewer undesirable side reactions (van den Burg *et al*, 2003). While hundreds of enzymes have been identified as potential biocatalysts, they are not always able to withstand industrial reaction conditions, which may require them to operate outside their optimal temperature range (Demirjian *et al*, 2001). The studies mentioned earlier, which have focussed on increasing the stability of enzymes, do not consider that a change in global stability will not guarantee an increase in high or low temperature activity, if T_{eq} remains unchanged. In contrast, the work presented here highlights the possibility of modifying existing enzymes which catalyse desirable reactions and tailoring them to suit the process temperature required for a particular application.

Chapter 7: Conclusions and future work

A combination of enzyme studies in the native organism coupled with recombinant characterisation of enzymes has provided strong support for the proposal made in chapter 1 for a novel pathway of D-xylose and L-arabinose metabolism in the hyperthermophilic archaeon *S. solfataricus.*

Labelling studies in the related organism *S. acidocaldarius* (P Schonheit, University of Kiel, Germany, personal communication) have shown that catabolism appears to be equally partitioned between this pathway and that proposed by Brouns *et al* (2006), although the factors regulating this partition are not yet known. It is known that all enzymes in the pathway proposed here are constitutively expressed in *S. solfataricus* and *S. acidocaldarius*, whereas enzymes from the alternative pathway are reported to be induced when the organism is grown on D-arabinose. It is however, important to remember that these organisms do not live in the nutrient rich environments to which they are exposed for the purposes of these studies. Perhaps the constitutively-expressed pathway described here represents how C5 sugars are metabolised in low nutrient environments, where the cost of inducing protein expression may be greater than the energy gained from catabolism. It is also possible that the organism has inducible pathways for when the substrate concentration in the environment is higher.

With respect to the individual enzymes characterised from this pathway; further investigation into the phosphorylation of the gluconate and xylonate dehydratases is required to provide insight into a possible regulation mechanism for a metabolic pathway comprised of constitutively-expressed enzymes. Full characterisation of glycolaldehyde oxidoreductase should also be undertaken, but was not possible during this project. While the data presented here indicate that the activity detected is that of a molybdenum containing aldehyde oxidoreductase, it is necessary to confirm the identity of the enzyme either through isolation and identification of the native protein or by producing a recombinant version from the putative genes already identified.

The molecular basis of the substrate promiscuity of glucose dehydrogenase proposed by Milburn *et al* (2006) was confirmed by the mutagenesis studies undertaken here. While the majority of the mutants constructed had detrimental effects on the substrate binding and

catalytic abilities of the enzyme, mutations of residue H297 did introduce additional ribose dehydrogenase activity with very little effect on the enzyme's natural function or thermal stability. Attempts were made to analyse these enzymes using the Equilibrium Model described in chapter 6, but due to the extreme thermostability of the enzymes involved, continuous assays could not be performed at temperatures high enough to collect data. It would, however, be extremely interesting to expand this study by introducing further mutations at this residue to see if ribose activity could be enhanced. By coupling this work with structural studies of the mutant enzymes, our understanding of the mechanism of promiscuity would be increased.

Studies on malate synthase provided strong evidence for the Equilibrium Model proposed by Daniel *et al* (2001). Determining the structure of malate synthase will be extremely important to understand the true positions and structural effects of these mutations, and may provide a basis for future mutagenesis studies on this enzyme.

Despite the importance of emerging *in silico* approaches, there is still a requirement to confirm enzyme function and metabolic interactions using isolated or recombinant proteins. An interesting addition to the *Sulfolobus* systems biology work, (Albers *et al*, 2009; Zaparty *et al* 2010) which looked at the effect of temperature on metabolism, would be a similar study incorporating a range of techniques to profile the gene and protein expression and analysis of metabolites in response to changes in carbohydrate type and availability in the growth environment of *Sulfolobus.* Coupled with the enzymological data presented here, a study such as this would provide a more complete picture of C5 metabolism, allowing for more accurate computation models to be produced in the future and enhance the understanding of archaeal metabolism.

References

Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS (2005) The 'evolvability' of promiscuous protein functions. Nature Genetics 37: 73-76.

Ahmed H, Tjaden B, Hensel R, Siebers B (2004) Embden-Meyerhof-Parnas and Entner-Doudoroff pathways in *Thermoproteus tenax*: metabolic parallelism or specific adaptation? Biochemical Society Transactions 32: 303-304.

Ahmed H, Ettema TJG, Tjaden B, Geerling ACM, van der Oost J, Siebers B (2005) The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. Biochemical Journal 390: 529-540.

Aivaliotis M, Macek B, Gnad F, Reichelt P, Mann M, Oesterhelt D (2009) Ser/Thr/Tyr Protein phosphorylation in the archaeon *Halobacterium salinarum* - A representative of the third domain of life. PLoS One 4: Article No: e4777.

Alber BE, Spanheimer R, Ebenau-Jehle C, Fuchs G (2006) Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. Molecular Microbiology 61: 297-309.

Albers SV, Elferink MGL, Charlebois RL, Sensen CW, Driessen AJM, Konings WN (1999) Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. Journal of Bacteriology 181: 4285-4291.

Albers SV, van de Vossenberg J, Driessen AJM, Konings WN (2000) Adaptations of the archaeal cell membrane to heat stress. Frontiers in Bioscience 5: D813-D820.

Albers SV, Van de Vossenberg J, Driessen AJM, Konings WN (2001) Bioenergetics and solute uptake under extreme conditions. Extremophiles 5: 285-294.

Albers SV, Jonuscheit M, Dinkelaker S, Urich T, Kletzin A, Tampe R, Driessen AJM, Schleper C (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Applied and Environmental Microbiology 72: 102-111.

Albers SV, Konings WN, Driessen AJM (2006) Membranes of thermophiles and other extremophiles. Extremophiles 35: 161-171.

Albers SV, Driessen AJM (2008) Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. Archaea 2: 1472-3646.

Albers SV, Driessen AJM (2008) Membranes and transport proteins of thermophilic microorganisms. Thermophiles: Biology and Technology at High Temperatures: 39-54.

Albers SV, Birkeland NK, Driessen AJM, Gertig S, Haferkamp P, Klenk HP, Kouril T, Manica A, Pham TK, Ruoff P et al. (2009) SulfoSYS (*Sulfolobus* Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. Biochemical Society Transactions 37: 58-64.

Alqueres SMC, Almeida RV, Clementino MM, Vieira RP, Almeida WI, Cardoso AM, Martins OB (2007) Exploring the biotechnological applications in the archaeal domain. Brazilian Journal of Microbiology 38: 398-405.

Antranikian G (2008) Extremozymes for industrial application. FEBS Journal 275: 48-48.

Arnold FH, Volkov AA (1999) Directed evolution of biocatalysts. Current Opinion in Chemical Biology 3: 54-59.

Assiddiq BF, Snijders APL, Chong PK, Wright PC, Dickman MJ (2008) Identification and characterization of *Sulfolobus solfataricus* P2 proteome using multidimensional liquid phase protein separations. Journal of Proteome Research 7: 2253-2261.

Babbitt PC, Hasson MS, Wedekind JE, Palmer DRJ, Barrett WC, Reed GH, Rayment I, Ringe D, Kenyon GL, Gerlt JA (1996) The enolase superfamily: A general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. Biochemistry 35: 16489-16501.

Babbitt PC, Gerlt JA (1997) Understanding enzyme superfamilies - Chemistry as the fundamental determinant in the evolution of new catalytic activities. Journal of Biological Chemistry 272: 30591-30594.

Baker PJ, Britton KL, Fisher M, Esclapez J, Pire C, Bonete MJ, Ferrer J, Rice DW (2009) Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. Proceedings of the National Academy of Sciences of the United States of America 106: 779-784.

Berglund P, Park S (2005) Strategies for altering enzyme reaction specificity for applied biocatalysis. Current Organic Chemistry 9: 325-336.

Bjelic S, Brandsdal BO, Aqvist J (2008) Cold adaptation of enzyme reaction rates. Biochemistry 47: 10049-10057.

Bjork A, Dalhus B, Mantzilas D, Sirevag R, Eijsink VGH (2004) Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface. Journal of Molecular Biology 341: 1215-1226.

Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. Proceedings of the National Academy of Sciences of the United States of America 103(15): 5869-5874.

Bloom JD, Arnold FH (2009) In the light of directed evolution: Pathways of adaptive protein evolution. Proceedings of the National Academy of Sciences of the United States of America 106: 9995-10000.

Bradford MM (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry 72: 248-254.

Brock TD, Brock KM, Belly RT, Weiss RL (1972) *Sulfolobus* - New genus of sulfur-oxidizing bacteria living at low pH and high-temperature. Archiv Fur Mikrobiologie 84: 54-&.

Brouns SJJ, Walther J, Snijders APL, de Werken H, Willemen H, Worm P, de Vos MGJ, Andersson A, Lundgren M, Mazon HFM et al. (2006) Identification of the missing links in prokaryotic pentose oxidation pathways - Evidence for enzyme recruitment. Journal of Biological Chemistry 281: 27378-27388.

Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiology and Molecular Biology Reviews 61: 456-+.

Buchanan CL, Connaris H, Danson MJ, Reeve CD, Hough DW (1999) An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. Biochemical Journal 343: 563-570.

Chen CY, Georgiev I, Anderson AC, Donald BR (2009) Computational structure-based redesign of enzyme activity. Proceedings of the National Academy of Sciences of the United States of America 106: 3764-3769.

Chothia C, Gough J (2009) Genomic and structural aspects of protein evolution. Biochemical Journal 419: 15-28.

Chung T, Klumpp DJ, Laporte DC (1988) Glyoxylate bypass operon of *Escherichia coli* - Cloning and determination of the functional map. Journal of Bacteriology 170: 386-392.

Ciaramella M, Pisani FM, Rossi M (2002) Molecular biology of extremophiles: recent progress on the hyperthermophilic archaeon *Sulfolobus*. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 81: 85-97.

Cirino PC, Georgescu R (2003) Screening for thermostability. Methods in Molecular Biology: 117-125.

Conway T (1992) The Entner-Doudoroff pathway: history, physiology and molecular biology. FEMS Microbiol Rev 9: 1-27.

Cornish-Bowden A, Eisenthal R (1978) Estimation of Michaelis constant and maximum velocity from Direct Linear plot. Biochimica Et Biophysica Acta 523: 268-272.

Dahms AS (1974) 3-Deoxy-D-Pentulosonic acid aldolase and Its role in a new pathway of D-xylose degradation. Biochemical and Biophysical Research Communications 60: 1433-1439.

D'Amico S, Marx JC, Gerday C, Feller G (2003) Activity-stability relationships in extremophilic enzymes. Journal of Biological Chemistry 278: 7891-7896.

Daniel RM, Peterson ME, Danson MJ, Price NC, Kelly SM, Monk CR, Weinberg CS, Oudshoorn ML, Lee CK. The molecular basis of the effect of temperature on enzyme activity. Biochemical Journal 425: 353-360.

Daniel RM, Danson MJ, Eisenthal R (2001) The temperature optima of enzymes: a new perspective on an old phenomenon. Trends in Biochemical Sciences 26: 223-225.

Daniel RM, Danson MJ, Eisenthal R, Lee CK, Peterson ME (2007) New parameters controlling the effect of temperature on enzyme activity. Biochemical Society Transactions 35: 1543-1546.

Daniel RM, Danson MJ, Eisenthal R, Lee CK, Peterson ME (2008) The effect of temperature on enzyme activity: new insights and their implications. Extremophiles 12: 51-59.

de Vos WM, Kengen SWM, Voorhorst WGB, van der Oost J (1998) Sugar utilization and its control in hyperthermophiles. Extremophiles 2: 201-205.

Demirjian DC, Moris-Varas F, Cassidy CS (2001) Enzymes from extremophiles. Current Opinion in Chemical Biology 5: 144-151.

Dengler U, Niefind K, Kiess M, Schomburg D (1997) Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*, NAD(+) and 2-oxoisocaproate at 1.9 Ångström resolution. Journal of Molecular Biology 267: 640-660.

DeRosa M, Gambacorta A, Nicolaus B, Giardina P, Poerio E, Buonocore V (1984) Glucose-metabolism in the extreme thermoacidophilic archaebacterium *Sulfolobus solfataricus*. Biochemical Journal 224: 407-414.

Doudoroff M (1940) The oxidative assimilation of sugars and related substances by *Pseudomonas saccharophila* with a contribution to the problem of the direct respiration of di- and polysaccharides. Enzymologia [Hague] 9: 59-72.

Dougherty MJ, Arnold FH (2009) Directed evolution: new parts and optimized function. Current Opinion in Biotechnology 20: 486-491.

Downs DM (2006) Understanding microbial metabolism. Annual Review of Microbiology 60: 533-559.

Driessen AJM, Szabo Z, Zolghadr B, Ellen AF, Ajon G, Albers SV (2008) Assembly and function of extracellular surface appendages and structures in the hyperthermoacidophile *Sulfolobus solfataricus.* FEBS Journal 275: 47-47.

Dumon-Seignovert L, Cariot G, Vuillard L (2004) The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expression and Purification 37: 203-206.

Edwards KJ, Barton JD, Rossjohn J, Thorn JM, Taylor GL, Ollis DL (1996) Structural and sequence comparisons of quinone oxidoreductase, zeta-crystallin, and glucose and alcohol dehydrogenases. Archives of Biochemistry and Biophysics 328: 173-183.

Eijsink VGH, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G (2004) Rational engineering of enzyme stability. Journal of Biotechnology 113: 105-120.

Eijsink VGH, Gaseidnes S, Borchert TV, van den Burg B (2005) Directed evolution of enzyme stability. Biomolecular Engineering 22: 21-30.

Eisenthal R, Cornish-Bowden A (1974) Direct linear plot - new graphical procedure for estimating enzyme kinetic-parameters. Biochemical Journal 139: 715-720.

Eisenthal R, Peterson ME, Daniel RM, Danson MJ (2006) The thermal behaviour of enzyme activity: implications for biotechnology. Trends in Biotechnology 24: 289-292.

Eisenthal R, Danson MJ, Hough DW (2007) Catalytic efficiency and k_{cat}/K_M: a useful comparator? Trends in Biotechnology 2: 247-249.

Elferink MGL, Albers SV, Konings WN, Driessen AJM (2001) Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. Molecular Microbiology 39: 1494-1503.

Ensign SA (2006) Revisiting the glyoxylate cycle: alternate pathways for microbial acetate assimilation. Molecular Microbiology 61: 274-276.

Entner N, Doudoroff M (1952) Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. Journal of Biological Chemistry 196: 853-862.

Falb M, Mueller K, Koenigsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D (2008) Metabolism of halophilic archaea. Extremophiles 12: 177-196.

Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH (2008) Evolutionary history of a specialized P450 propane monooxygenase. Journal of Molecular Biology 383: 1069-1080.

Fauvart M, Braeken K, Daniels R, Vos K, Ndayizeye M, Noben JP, Robben J, Vanderleyden J, Michiels J (2007) Identification of a novel glyoxylate reductase supports phylogeny-based enzymatic substrate specificity prediction. Biochimica Et Biophysica Acta-Proteins and Proteomics 1774: 1092-1098.

Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN, Timmis KN (2003) Chaperonins govern growth of *Escherichia coli* at low temperatures. Nature Biotechnology 21: 1266-1267.

Gaseidnes S, Synstad B, Nielsen JE, Eijsink VGH (2003) Rational engineering of the stability and the catalytic performance of enzymes. Journal of Molecular Catalysis B-Enzymatic 21: 3-8.

Georlette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, Marx JC, Sonan G, Feller G, Gerday C (2004) Some like it cold: biocatalysis at low temperatures. FEMS Microbiology Reviews 28: 25-42.

Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. Annual Review of Biochemistry 70: 209-246.

Gerlt JA, Raushel FM (2003) Evolution of function in (beta/alpha) (8)-barrel enzymes. Current Opinion in Chemical Biology 7: 252-264.

Gerlt JA (2004) How enzymes evolve: Insights from natural and unnatural promiscuity. Abstracts of Papers American Chemical Society 228: U181.

Gerlt JA, Babbitt PC, Rayment I (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Archives of Biochemistry and Biophysics 433: 59-70.

Gerlt JA, Babbitt PC (2009) Enzyme (re)design: lessons from natural evolution and computation. Current Opinion in Chemical Biology 13: 10-18.

Giver L, Gershenson A, Freskgard PO, Arnold FH (1998) Directed evolution of a thermostable esterase. Proceedings of the National Academy of Sciences of the United States of America 95: 12809-12813.

Glasner ME, Gerlt JA, Babbitt PC (2006) Evolution of enzyme superfamilies. Current Opinion in Chemical Biology 10: 492-497.

Godoy-Ruiz R, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM (2004) Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. Journal of Molecular Biology 336: 313-318.

Goldberg JD, Yoshida T, Brick P (1994) Crystal structure of a NAD dependent D-glycerate dehydrogenase at 2.4 Ångström resolution. Journal of Molecular Biology 236: 1123-1140.

Gomes J, Steiner W (2004) The biocatalytic potential of extremophiles and extremozymes. Food Technology and Biotechnology 42: 223-235.

Graslund S, Sagemark J, Berglund H, Dahlgren LG, Flores A, Hammarstroem M, Johansson I, Kotenyova T, Nilsson M, Nordlund P et al. (2008) The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. Protein Expression and Purification 58: 210-221.

Grogan DW (1989) Phenotypic characterization of the archaebacterial genus *Sulfolobus* - comparison of 5 wild-type strains. Journal of Bacteriology 171: 6710-6719.

Hain J, Reiter WD, Hudepohl U, Zillig W (1992) Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Research 20: 5423-5428.

Hasson MS, Schlichting I, Moulai J, Taylor K, Barrett W, Kenyon GL, Babbitt PC, Gerlt JA, Petsko GA, Ringe D (1998) Evolution of an enzyme active site: The structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase. Proceedings of the National Academy of Sciences of the United States of America 95: 10396-10401.

Heinzelman P, Snow CD, Smith MA, Yu XL, Kannan A, Boulware K, Villalobos A, Govindarajan S, Minshull J, Arnold FH (2009) SCHEMA Recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. Journal of Biological Chemistry 28: 26229-26233.

Hough DW, Danson MJ (1999) Extremozymes. Current Opinion in Chemical Biology 3: 39-46.

Howard BR, Endrizzi JA, Remington SJ (2000) Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Ångström resolution: Mechanistic implications. Biochemistry 39: 3156-3168.

Hu YJ, Holden JF (2006) Citric acid cycle in the hyperthermophilic archaeon *Pyrobaculum islandicum* grown autotrophically, heterotrophically, and mixotrophically with acetate. Journal of Bacteriology 188: 4350-4355.

Huegler M, Huber H, Stetter KO, Fuchs G (2003) Autotrophic CO2 fixation pathways in archaea (Crenarchaeota). Archives of Microbiology 179: 160-173.

Hurwitz J, Weissbach A (1959) Formation of 2-keto-3-deoxyheptonic acid in extracts of *Escherichia coli* B2. Enzymic studies. Journal of Biological Chemistry 234(4): 710-712.

Jensen RA (1976) Enzyme recruitment in evolution of new function. Annual Review of Microbiology 30: 409-425.

Jiang D, Zhou SG, Chen YPP (2009) Compensatory ability to null mutation in metabolic networks. Biotechnology and Bioengineering 103: 361-369.

Johannes TW, Zhao HM (2006) Directed evolution of enzymes and biosynthetic pathways. Current Opinion in Microbiology 9: 261-267.

Johnsen U, Dambeck M, Zaiss H, Fuhrer T, Soppa J, Sauer U, Schoenheit P (2009) D-Xylose degradation pathway in the halophilic archaeon *Haloferax volcanii*. Journal of Biological Chemistry 284: 27290-27303.

Johnson KA (2008) Role of induced fit in enzyme specificity: A molecular forward/reverse switch. Journal of Biological Chemistry 283: 26297-26301.

Kardinahl S, Schmidt CL, Hansen T, Anemuller S, Petersen A, Schafer G (1999) The strict molybdate-dependence of glucose-degradation by the thermoacidophile *Sulfolobus acidocaldarius* reveals the first crenarchaeotic molybdenum containing enzyme - an aldehyde oxidoreductase. European Journal of Biochemistry 260: 540-548.

Kennelly PJ (2001) Protein phosphatases - A phylogenetic perspective. Chemical Reviews 101(8): 2291-2312.

Kennelly PJ (2002) Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. FEMS Microbiology Letters 206: 1-8.

Kennelly PJ (2003) Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry. Biochemical Journal 370: 373-389.

Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Current Opinion in Chemical Biology 10: 498-508.

Kim SH, Lee SB (2005) Identification and characterization of *Sulfolobus solfataricus* D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner-Doudoroff pathway. Biochemical Journal 387: 271-280.

Kim S, Lee SB (2006) Rare codon clusters at 5 '-end influence heterologous expression of archaeal gene in *Escherichia coli*. Protein Expression and Purification 50: 49-57.

Kim S, Lee SB (2006) Catalytic promiscuity in dihydroxy-acid dehydratase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Journal of Biochemistry 139(3): 591-596.

Kim S, Lee SB (2008) Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. Protein Expression and Purification 62: 116-119.

Koma D, Sawai T, Harayama S, Kino K (2006) Overexpression of the genes from thermophiles in *Escherichia coli* by high-temperature cultivation. Applied Microbiology and Biotechnology 73: 172-180.

Konings WN, Albers SV, Koning S, Driessen AJM (2002) The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 81: 61-72.

Kornberg HL, Krebs HA (1957) Synthesis of Cell Constituents from C2-Units by a Modified Tricarboxylic Acid Cycle. Nature 179: 988-991.

Lamble HJ, Heyer NI, Bull SD, Hough DW, Danson MJ (2003) Metabolic pathway promiscuity in the Archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. Journal of Biological Chemistry 278: 34066-34072.

Lamble FJ, Milburn CC, Taylor GL, Hough DW, Danson MJ (2004) Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in *Sulfolobus solfataricus*. FEBS Letters 576: 133-136.

Lamble HJ, Theodossis A, Milburn CC, Taylor GL, Bull SD, Hough DW, Danson MJ (2005) Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. FEBS Letters 579: 6865-6869.

Leblanc DJ, Mortlock RP (1971) Metabolism of D-arabinose - New pathway in *Escherichia coli.* Journal of Bacteriology 106: 90-&.

Lee CK, Daniel RM, Shepherd C, Saul D, Cary SC, Danson MJ, Eisenthal R, Peterson ME (2007) Eurythermalism and the temperature dependence of enzyme activity. FASEB Journal 21: 1934-1941.

Lee SC, Chang Y, Shin D-M, Han J, Seo M-H, Fazelinia H, Maranas CD, Kim H-S (2009) Designing the substrate specificity of D-hydantoinase using a rational approach. Enzyme and Microbial Technology 44: 170-175.

Lehmann M, Wyss M (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Current Opinion in Biotechnology 12: 371-375.

Lewis JC, Arnold FH (2009) Catalysts on semand: Selective oxidations by laboratory-evolved cytochrome P450 BM3. Chimia 63: 309-312.

Lohman JR, Olson AC, Remington SJ (2008) Atomic resolution structures of *Escherichia coli* and *Bacillus anthracis* malate synthase A: Comparison with isoform G and implications for structure-based drug discovery. Protein Science 17: 1935-1945.

Lower BH, Bischoff KM, Kennelly PJ (2000) The archaeon *Sulfolobus solfataricus* contains a membrane-associated protein kinase activity that preferentially phosphorylates threonine residues in vitro. Journal of Bacteriology 182: 3452-3459.

Lower BH, Kennelly PJ (2003) Open reading frame sso2387 from the archaeon *Sulfolobus solfataricus* encodes a polypeptide with protein-serine kinase activity. Journal of Bacteriology 185: 3436-3445.

Lower BH, Ben Potters M, Kennelly PJ (2004) A phosphoprotein from the archaeon *Sulfolobus solfataricus* with protein-serine/threonine kinase activity. Journal of Bacteriology 186: 463-472.

Lubelska JM, Jonuscheit M, Schleper C, Albers SV, Driessen AJM (2006) Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon *Sulfolobus solfataricus*. Extremophiles 10: 383-391.

Luke KA, Higgins CL, Wittung-Stafshedel P (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. FEBS Journal 274: 4023-4033.

Macgee J, Doudoroff M (1954) A new phosphorylated intermediate in glucose oxidation. Journal of Biological Chemistry 210: 617-626.

Mathias AL, Rigo LU, Funayama S, Pedrosa FO (1989) L-Arabinose metabolism in *Herbaspirillum seropedicae*. Journal of Bacteriology 171: 5206-5209.

Mayr E (1990) A Natural System of Organisms. Nature 348(6301): 491-491.

Milburn CC, Lamble HJ, Theodossis A, Bull SD, Hough DW, Danson MJ, Taylor GL (2006) The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. Journal of Biological Chemistry 281: 14796-14804.

Miyazaki K, Arnold FH (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. Journal of Molecular Evolution 49: 716-720.

Moll R, Schafer G (1988) Chemiosmotic-H+ cycling across the plasma-membrane of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Letters 232: 359-363.

Nath A, Atkins WM (2008) A quantitative index of substrate promiscuity. Biochemistry 47: 157-166.

Nobeli I, Favia AD, Thornton JM (2009) Protein promiscuity and its implications for biotechnology. Nature Biotechnology 27: 157-167.

O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. Chemistry & Biology 6: R91-R105.

Ogino H, Nakayama H, China H, Kawata T, Doukyu N, Yasuda M (2008) Characterization of recombinant glyoxylate reductase from thermophile *Thermus thermophilus* HB27. Biotechnology Progress 24: 321-325.

Ohshima T, Nunoura-Kominato N, Kudome T, Sakuraba H (2001) A novel hyperthermophilic archaeal glyoxylate reductase from *Thermococcus litoralis* - Characterization, gene cloning, nucleotide sequence and expression in Escherichia coli. European Journal of Biochemistry 268: 4740-4747.

Pedrosa FO, Zancan GT (1974) L-Arabinose metabolism in *Rhizobium japonicum*. Journal of Bacteriology 119: 336-338.

Peeters E, Albers SV, Vassart A, Driessen AJM, Charlier D (2009) Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes. Molecular Microbiology 71: 972-988.

Peterson ME, Eisenthal R, Danson MJ, Spence A, Daniel RM (2004) A new intrinsic thermal parameter for enzymes reveals true temperature optima. Journal of Biological Chemistry 279: 20717-20722.

Peterson ME, Eisenthal R, Danson MJ, Spence A, Daniel RM (2005) A new intrinsic thermal parameter for enzymes reveals true temperature optima. Journal of Biological Chemistry 280: 41784-41784.

Peterson ME, Daniel RM, Danson MJ, Eisenthal R (2007) The dependence of enzyme activity on temperature: determination and validation of parameters. Biochemical Journal 402: 331-337.

Pire C, Esclapez J, Diaz S, Perez-Pomares F, Ferrer J, Bonete MJ (2009) Alteration of coenzyme specificity in halophilic NAD(P)$^+$ glucose dehydrogenase by site-directed mutagenesis. Journal of Molecular Catalysis B-Enzymatic 59: 261-265.

Potter JA, Kerou M, Lamble HJ, Bull SD, Hough DW, Danson MJ, Taylor GL (2008) The structure of *Sulfolobus solfataricus* 2-keto-3-deoxygluconate kinase. Acta Crystallographica Section D-Biological Crystallography 64: 1283-1287.

Potters BM, Solow BT, Bischoff KM, Graham DE, Lower BH, Helm R, Kennelly PJ (2003) Phosphoprotein with phosphoglycerate mutase activity from the archaeon *Sulfolobus solfataricus*. Journal of Bacteriology 185: 2112-2121.

Prangishvili D, Albers SV, Holz I, Arnold HP, Stedman K, Klein T, Singh H, Hiort J, Schweier A, Kristjansson JK et al. (1998) Conjugation in archaea: Frequent occurrence of conjugative plasmids in *Sulfolobus*. Plasmid 40: 190-202.

Rakus JF, Fedorov AA, Fedorov EV, Glasner ME, Hubbard BK, Delli JD, Babbitt PC, Almo SC, Gerlt JA (2008) Evolution of enzymatic activities in the enolase superfamily: L-rhamnonate dehydratase. Biochemistry 47: 9944-9954.

Rakus JF, Kalyanaraman C, Fedorov AA, Fedorov EV, Mills-Groninger FP, Toro R, Bonanno J, Bain K, Sauder JM, Burley SK et al. (2009) Computation-facilitated assignment of the function in the enolase superfamily: A regiochemically distinct galactarate dehydratase from *Oceanobacillus iheyensis*. Biochemistry 48:11546-11558.

Razeto A, Kochhar S, Hottinger H, Dauter M, Wilson KS, Lamzin VS (2002) Domain closure, substrate specificity and catalysis of D-Lactate dehydrogenase from *Lactobacillus bulgaricus*. Journal of Molecular Biology 318: 109-119.

Reiter WD, Hudepohl U, Zillig W (1990) Mutational analysis of an archaebacterial promoter - Essential role of a TATA Box for transcription efficiency and start-site selection *in vitro*. Proceedings of the National Academy of Sciences of the United States of America 87: 9509-9513.

Roca M, Liu H, Messer B, Warshel A (2007) On the relationship between thermal stability and catalytic power of enzymes. Biochemistry 46: 15076-15088.

Romano AH, Conway T (1996) Evolution of carbohydrate metabolic pathways. Research in Microbiology 147: 448-455.

Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. Nature Reviews Molecular Cell Biology 10: 866-876.

Ronimus RS, Morgan HW (2003) Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. Archaea 1: 199-221.

Rubin-Pitel SB, Zhao HM (2006) Recent advances in biocatalysis by directed enzyme evolution. Combinatorial Chemistry & High Throughput Screening 9: 247-257.

Sacchi S, Lorenzi S, Molla G, Pilone MS, Rossetti C, Pollegioni L (2002) Engineering the substrate specificity of D-amino-acid oxidase. Journal of Biological Chemistry 277: 27510-27516.

Saito N, Robert M, Kitamura S, Baran R, Soga T, Mori H, Nishioka T, Tomita M (2006) Metabolomics approach for enzyme discovery. Journal of Proteome Research 5: 1979-1987.

Sakai A, Fedorov AA, Fedorov EV, Schnoes AM, Glasner ME, Brown S, Rutter ME, Bain K, Chang S, Gheyi T et al. (2009) Evolution of enzymatic activities in the enolase superfamily: Stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes. Biochemistry 48: 1445-1453.

Sawayama AM, Chen MMY, Kulanthaivel P, Kuo MS, Hemmerle H, Arnold FH (2009) A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. Chemistry European Journal 15: 11723-11729.

Schmid AK, Reiss DJ, Pan M, Koide T, Baliga NS (2009) A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. Molecular Systems Biology 5:282.

Schmidt DMZ, Gerlt JA (2002) The enolase superfamily: assigning function and evolving activities. Faseb Journal 16: A904-A904.

Schmidt DMZ, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA (2003) Evolutionary potential of (beta/alpha)(8)-barrels: Functional promiscuity produced by single substitutions in the enolase superfamily. Biochemistry 42: 8387-8393.

Schonheit P, Schafer T (1995) Metabolism of hyperthermophiles. World Journal of Microbiology & Biotechnology 11: 26-57.

Schuller DJ, Grant GA, Banaszak LJ (1995) The allosteric ligand site in the V-max-type cooperative enzyme phosphoglycerate dehydrogenase. Nature Structural Biology 2: 69-76.

Scopes RK (1995) The effect of temperature on enzymes used in diagnostics. Clinica Chimica Acta 237: 17-23.

Selig M, Xavier KB, Santos H, Schonheit P (1997) Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. Archives of Microbiology 167: 217-232.

Sen S, Dasu VV, Mandal B (2007) Developments in directed evolution for improving enzyme functions. Applied Biochemistry and Biotechnology 143: 212-223.

Serrano L, Day AG, Fersht AR (1993) Step-wise mutation of barnase to binase - a procedure for engineering increased stability of proteins and an experimental-analysis of the evolution of protein stability. Journal of Molecular Biology 233: 305-312.

Serrano JA, Camacho M, Bonete MJ (1998) Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in *Haloferax volcanii*. FEBS Letters 434: 13-16.

Serrano JA, Bonete MJ (2001) Sequencing, phylogenetic and transcriptional analysis of the glyoxylate bypass operon (ace) in the halophilic archaeon *Haloferax volcanii*. Biochimica Et Biophysica Acta-Gene Structure and Expression 1520: 154-162.

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CCY, Clausen IG, Curtis BA, De Moors A et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proceedings of the National Academy of Sciences of the United States of America 98: 7835-7840.

Siebers B, Schonheit P (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Current Opinion in Microbiology 8: 695-705.

Spiller B, Gershenson A, Arnold FH, Stevens RC (1999) A structural view of evolutionary divergence. Proceedings of the National Academy of Sciences of the United States of America 96: 12305-12310.

Stetter KO (2006) History of discovery of the first hyperthermophiles. Extremophiles 10: 357-362.

Sunnarborg A, Klumpp D, Chung T, Laporte DC (1990) Regulation of the glyoxylate bypass operon - cloning and characterization of Iclr. Journal of Bacteriology 172: 2642-2649.

Theodossis A, Walden H, Westwick EJ, Connaris H, Lamble HJ, Hough DW, Danson MJ, Taylor GL (2004) The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus*. Journal of Biological Chemistry 279: 43886-43892.

Theodossis A, Milburn CC, Heyer NI, Lamble HJ, Hough DW, Danson MJ, Taylor G (2005) Preliminary crystallographic studies of glucose dehydrogenase from the promiscuous Entner-Doudoroff pathway in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Acta Crystallographica Section F-Structural Biology and Crystallization Communications 61: 112-115.

Thomas TM, Scopes RK (1998) The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. Biochemical Journal 330: 1087-1095.

Tokuriki N, Stricher F, Serrano L, Tawfik DS (2008) How Protein Stability and New Functions Trade Off. Plos Computational Biology 4(2).

Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. Current Opinion in Structural Biology 19: 596-604.

Tolstrup N, Sensen CW, Garrett RA, Clausen IG (2000) Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. Extremophiles 4: 175-179.

Trevisanato SI, Larsen N, Segerer AH, Stetter KO, Garrett RA (1996) Phylogenetic analysis of the archaeal order of *Sulfolobales* based on sequences of 23S rRNA genes and 16S/23S rDNA spacers. Systematic and Applied Microbiology 19: 61-65.

Uhrigshardt H, Walden M, John H, Petersen A, Anemuller S (2002) Evidence for an operative glyoxylate cycle in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. FEBS Letters 513: 223-229.

Unsworth LD, van der Oost J, Koutsopoulos S (2007) Hyperthermophilic enzymes - stability, activity and implementation strategies for high temperature applications. FEBS Journal 274: 4044-4056.

Van den Burg B, Vriend G, Veltman OR, Eijsink VGH (1998) Engineering an enzyme to resist boiling. Proceedings of the National Academy of Sciences of the United States of America 95: 2056-2060.

Van den Burg B, Eijsink VGH (2002) Selection of mutations for increased protein stability. Current Opinion in Biotechnology 13: 333-337.

Van den Burg B (2003) Extremophiles as a source for novel enzymes. Current Opinion in Microbiology 6: 213-218.

Verhees CH, Kengen SWM, Tuininga JE, Schut GJ, Adams MWW, De Vos WM, Van der Oost J (2003) The unique features of glycolytic pathways in Archaea. Biochemical Journal 375: 231-246.

Vick JE, Gerlt JA (2007) Evolutionary potential of (beta/alpha) (8)-Barrels: Stepwise evolution of a "New" reaction in the enolase superfamily. Biochemistry 46: 14589-14597.

Wagner M, Berkner S, Ajon M, Driessen AJM, Lipps G, Albers SV (2009) Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. Biochemical Society Transactions 37: 97-101.

Wang B, Yang SF, Zhang L, He ZG (2010) Archaeal eukaryote-like serine/threonine protein kinase interacts with and phosphorylates a forkhead-associated-domain-containing protein. Journal of Bacteriology 192: 1956-1964.

Watanabe S, Shimada N, Tajima K, Kodaki T, Makino K (2006) Identification and characterization of L-arabonate dehydratase, L-2-keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism - Novel evolutionary insight into sugar metabolism. Journal of Biological Chemistry 281: 33521-33536.

Watanabe S, Saimura M, Makino K (2008) Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. Journal of Biological Chemistry 283: 20372-20382.

Watanabe S, Piyanart S, Makino K (2008) Metabolic fate of L-lactaldehyde derived from an alternative L-rhamnose pathway. FEBS Journal 275: 5139-5149.

Watanabe S, Makino K (2009) Novel modified version of nonphosphorylated sugar metabolism - an alternative L-rhamnose pathway of *Sphingomonas* sp. FEBS Journal 276: 1554-1567.

Weinberg CS, Daniel RM, Monk CR, Danson MJ, Lee CK (2008) The Equilibrium Model for the effect of temperature on enzymes - Insights and implications. Chimica Oggi-Chemistry Today 26: 14-15.

Weissbach A, Hurwitz J (1959) Formation of 2-keto-3-deoxyheptonic acid in extracts of *Escherichia coli* B1. Identification. Journal of Biological Chemistry 234: 705-709.

Wintrode PL, Zhang DQ, Vaidehi N, Arnold FH, Goddard WA (2003) Protein dynamics in a family of laboratory evolved thermophilic enzymes. Journal of Molecular Biology 327: 745-757.

Woese CR (1979) Proposal concerning the origin of life on the planet earth. Journal of Molecular Evolution 13: 95-101.

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms - Proposal for the domains archaea, bacteria, and eucarya. Proceedings of the National Academy of Sciences of the United States of America 87: 4576-4579.

Wolfe AJ (2005) The acetate switch. Microbiology and Molecular Biology Reviews 69: 12-50.

Wolterink-van Loo S, van Eerde A, Siemerink MAJ, Akerboom J, Dijkstra BW, van der Oost J (2007) Biochemical and structural exploration of the catalytic capacity of *Sulfolobus* KDG aldolases. Biochemical Journal 403: 421-430.

Yamamoto K, Ishihama A (2003) Two different modes of transcription repression of the *Escherichia coli* acetate operon by IcIR. Molecular Microbiology 47: 183-194.

Yoshikawa S, Arai R, Kinoshita Y, Uchikubo-Kamo T, Wakamatsu T, Akasaka R, Masui R, Terada T, Kuramitsu S, Shirouzu M et al. (2007) Structure of archaeal glyoxylate reductase from Pyrococcus *horikoshii* OT3 complexed with nicotinamide adenine dinucleotide phosphate. Acta Crystallographica Section D-Biological Crystallography 63: 357-365.

Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440: 1078-1082.

Zaparty M, Tjaden B, Hensel R, Siebers B (2008) The central carbohydrate metabolism of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: pathways and insights into their regulation. Archives of Microbiology 190: 231-245.

Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P et al (2010). "Hot standards" for the thermoacidophilic archaeon *Sulfolobus solfataricus*. Extremophiles 14: 119-142.

Zhao HM, Arnold FH (1999) Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Engineering 12: 47-53.

Zillig W, Palm P, Reiter WD, Gropp F, Puhler G, Klenk HP (1988) Comparative-evaluation of gene-expression in archaebacteria. European Journal of Biochemistry 173: 473-482.

Zillig W, Prangishvilli D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D, Gotz D (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. FEMS Microbiology Reviews 18(2-3): 225-236.

Appendix 1: Sequence of genes associated with a novel metabolic pathway for C5 sugars in *S. solfataricus*.

Sequence of genes (uppercase) and upstream sequences (lowercase) are shown, highlighting putative TATA boxes in the case of single genes (Zillig, 1988) and a Shine-Dalgarno sequence for glycolaldehyde oxidoreductase genes located inside an operon (Trevisanato, 1996). These putative transcription initiation sites are underlined in both cases.

1a. Xylonate dehydratase

tattgtgtagccgataaaaactttatttttccaaaccttacttacatacATGACAAAGATTTCAGAA
ATCGAAGCGTACATTTTAGGTAAGGAAGTAACAAGCGCTCAATGGGCTTCATTAATGGTACTGGTAAG
GGTAACCACAAACGATGGTAGAGTAGGCTGGGGAGAAACTGTAAGTGCCTTAAGGGCAGAAGCTGTGG
CAAATTTTGTAAAAAAGATAAATACTGTATTAAAAGGAAATGACGTGTTTAACGTTGAAAAGAACAGG
CTAGAATGGTATAAACATGATTTCAATATGACTATCTCCTTAGAGTCTACAACAGCCTATAGTGCAGT
AGATATAGCTTCTTGGGACATAATCGGAAAAGAGTTAGGTGCTCCATTATATAAGCTGCTTGGAGGAA
AAACAAGAGATAAAGTACTTGTCTACGCTAATGGTTGGTATCAAAACTGCGTGAAGCCAGAGGACTTT
GCGGAAAAAGCTAAGGAGATCGTAAAAATGGGATATAAGGCTTTGAAATTTGACCCATTTGGTCCGTA
CTTTAACGATATTTCAAAGAAAGGACTAGATATAGCTGAGGAGAGAGTAAAGGCTGTTAGAGAGGCAG
TTGGAGACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTCGGCTATTATGATA
GCGAAAAGACTAGAGAAGTATAATCCATTGTTCATGGAAGAACCAATTCATCCTGAGGATGTGGAAGG
ACTTAGAAAATATAGGAACAATACGAGCTTAAGGATTGCGTTAGGTGAAAGGATAATTAATAAACAGC
AAGCCTTATACTTTATGAAAGAGGGATTAGTGGATTTCTTACAAGCTGATCTATACAGAATTGGTGGA
GTTACTGAGACCAAGAAGGTAGTAGGGATTGCTGAGACCTTTGACGTACAGATGGCTTTTCACAATGC
TCAAGGTCCAATACTAAACGCCGTAACACTACAGTTTGACGCATTTATACCAAACTTCTTAATACAAG
AGTCTTTCTATGATTGGTTCCCGAGTTGGAAGAGGGAGCTAATATATAATGGTACTCCAATAGATAAC
GGATACGCTATAATACCGGAAAGACCGGGTTTAGGAGTTGAAGTAAATGAGAAAATGTTAGATAGTTT
AAAGGTTAAAGGTGAGGAGTACTTTAATCCAGAAGAACCAGTGTGGGTAGTTAAAGGAACTTGGAGAG
ACTATTAG

1b. Malate synthase / Isocitrate lyase

aggtactccatatatctaattgaaacttttttagaaaaatttaatagtattctccaacttttttaaaat
tggggatcaaATGTCTTCTAGCCTTAAAATCCCTGAGGAGATCTATCAAAAGTATAGTGACCTATTCG
GAGAAAAGACCATAAACGATAGGATTGTAAGTGTTGAAAAACTAATTGAGGAGCTCGCAGTTGAGTTC
TCAGATGAGATAAGAAAGGTTATTAATAAAAGGCGACTATGGTTAGAATCAAAAGACCCCGTGACCTC
AAAAGGTGCTTTTCCATCCTTTGACGAAGTCTTTGTAGATGCTGATGGAAATAAGAGAACCTTTAGGG
AAATCATTCAAGGGATGATAGATAACTTCCTCGGGGTACAGTCAAAGCTAAAATGGAGGTTAAATGAA
AATGTACCAATACCAAAGGACGCACATCCCCTCAATAACCCAGGGTTAGAAATCACTGGACCTTGGTA
TCCTCTGAGTAGGGCGTATAATCAAATAAACTCAGACGTAGCATGTGTTATGGAAGATGAGGAAGACG
CCTCACCTGCATGGTATATACCCTTTGGTTCTGGTAAAACAATTGCTGACGTATGGGAGGGAAGAAAG
AACGTAAAACTCTTCCTCTCCGGTAAGGCTCCAAATCCTTATTACGAGAAGGGAAAAACGTACAGCTT
AAACAAACCTAGGGATAAGTGGCCAGTGATTTTCCACAGACTCCCTGGATTACACCTCTTGGATTTCG
ATATAACGTTAAATGGTAAACCAGTACCGGCTATAATTGTTTCTGCCGTTATATATACCCTAAACAAT
TATAATAGTCTGAAGAGTGCGGGTTCTGGAGTATATTTCTATCTACCTAAAACCCAGACACCGGATGA
GGCATTGATAATAGAAAAGATCTTGAGAAGAATCGAGTCTAAACTAGGACTTAAGATAGGTACACTTA
AGCTCGCTCTCCTTTATGAGGAGGTCAATGCTGGAAGATTTTTCCCTGTAATACTGTGGATATTTCGT
GAGAGACTCATAAAATCAAATAACGGTAGATGGGACTATTTGGGAAGTCTGATTGAGATGTGGTTACA
AGAGAAAGTACTTCCAGATCCCCAAAACATTACCATGACCTCACCAAACATGATGGCTTATCAAAAGT
ACAACGCACTCATGATGCTATTAGCTGGGGCGAAGAACGGCGAAGCCGATGCTGCTCCAGTAGGTGGT
ATGGCGGCAGTAATGCTCTACCCTCAAACTGACCCATTCGGGAGAAATAGATACAACTTAAAGGCATT
AAGGGGAATAAAGCTAGATAAGTTAAGGGAGAGGCTAATAGGACTTATCTTTGTGGCAGAGGGTAAAG
TTGAAGGTAAAATCACTTTAGAGGATATAGTTAACGGAAAAGTAAAGGGTAAACTTTACGATATGTTC
AGACAGAGTTGGGTAGCGACTAAGGAAGAAGCTTATGTAGAGGCAGGGAGCAAACCCTTAAGGGCAAG
TTTAGATGAGTTACAGAAAATGATTGACGCACCAATTAATTACATTGAGGTAGAGGGAACCAAATTAC

CCACCGTGGATAGTGGGTTAACACCAGAAGAAAGAGCGTTATTCCAAAAGCTTGGCCTAATTGATGAA
AGAGGAAAAATAACGCCTTGGGTTATAACCAAGGAAATGATTAACACTCCAGAAAAGCTATTATTTAA
TAAGGAATTATGGGGTGGAAAAGACCTATGGCATTCGTTATATGACATCCCAGAGGGAGACATTACAC
CAGAACACGTTCAACACGCGTTTTATATGGCAGCGAATTACGGTTTCCAATTGCTAAATGGCAACTTA
GCTGCGGCAATAGATGATTATGAATTAAAGCAAAGGTTCATGAACGACTTAGCTACCTATAGAATATT
CACTTCTTGGCTATGGAGTGTAATAAATAGGGATGCTAGCTTCACAAAGGATGGATATATTAAGGGAC
CAAAGCTAACTAAGGATGGTGTAATTCCAGCTGAAGACGTTCTAAAGGTCACAAAGGGGACGAAGGTA
AAGGATATATTCGAGAAGTTGTGGGAGTTACACTTAGATTGGACATATGAGTTTTACAAAGAACAAGA
TATGAGAGCGTCTAGGAGAATAGCTGAAACCTTTGGAAAGACAAATAACGCTTCAACGGTGGAAGAGG
TGTATAAGGTAATATCTAAGGCTTATAGTTCAGGTCCGTTTAGGGAGATGTCTGCTAAAGAAGCTGCT
CAGAAGCTGGCAAAAATCCTTAACGCTAACGCTTCTGAGATCGAGGAAGAGCTTATTAATTTGGCCCC
GAGGTTTGACAGATCAATGGCCCCAGTAATTATGGAAATATTAATGAAAGAGATGTTATATCCCAAGT
ACATAATGAATAGCGGGAAGATACTCTTCATACTCTCTCCATTGGATCCAGAGAGGAGGTCAAAGGTA
ATGGATAGCATATTCTCGTTTAGAAAAATGATTGAAGATAAGGTAAAAAGAGGAGAACTGGATAAGTG
GGTATTGGAGTTATACGACTACGTTTACGACAATTACTGGTAA**ataggttaattcttattatttttta**
aacaaaaaatgaaggggactcttctttttgtaaaaaaatcgtaatttacataaggttatctcgttct
actgccttgatctaaatacttttattttctctaatttttatgttaagtaaacgcatagctaaaggct
tatgcgagttatagctagagtttcgttaaaaagtgtttagttttctcaattagagtcaagtaaaattc
tactaatgacgttgaagatctaactataatttcgctttttaggacgtattaaaagatttctttttgtc
aatattacatggaaatgttgactataaattccttaaaatataaatacttttttagaccttgccaagaag
ttgataaaatatctttaaattggatagggaaagacataaactattaatgctaaagtttatata**ttaaa**
atatgtgtatatagattaggatgaaaaaATGAACATCCGAGATAAATGGATTGAAGAGGAGAAAAGTC
TCGAATATGAGTGGTCTGAAGATCCAAGGTGGAGAGGAATAAAGAGAAATTATAAGCCCTTGGATGTG
GTTAAATTAAGAGGATCTATTAGAATTGAGTATAGTTTAGCTAAGTTAGCCTCTCACAAACTTTGGAA
CTTACTAAATACTGAGCCCTATGTGGCAACATTTGGGGCATTAACGGGTTCCCAAGCAGTGGAAATGG
CAAAGGCGGGCCTTAAGGCTATTTACGTAAGTGGATGGCAAGTAGCTGCAGATAACAACTTATCAAAT
CAAACTTATCCAGACTTAAGCTTATACCCATCTAACAGCGTTCCAAACCTAGTAAAGAGGTTAAATAA
CGCCCTTATAAGGGCTGATCAAATATCTTGGAGTGAGGGTAAACATGATATTGACTATTTACTGCCTA
TTGTTGCAGATGCTGAAGCTGGTTTTGGTGGTCCAATTCACGCCTTCGAATTAACAAAAGCCTTAATA
GAGGCAGGGGCTGCAGGGGTACATTTCGAAGATCAATTAGCAGCAGAGAAGAAGTGTGGCCATTTAGG
AGGGAAAGTTCTGATCCCTATAAGTGCATTTATAAGAGTACTAAACGCTGCAAGACTAGCGTCTGACG
TATTAGGAGTACCAACAATACTAATAGCAAGGACAGACGCACTGAATGCTAAATACTTATCTAACGAT
GTAGATGAGACTGATAATCAATTCTTAACAGGAAAGAGGACATCAGAGGGATATTATGAGATAAAGGG
TGGTATTGAATACGCCATAGCTAGAGGATTAGCTTATGCACCCTATGCTGACCTATTGTGGTTTGAGA
CTTCTAAGCCAGATTTGGAAGAGGCTAGACAATTTGCTGAGGCTATTCACACTCATTATCCAGGTAAG
ATGTTAGCGTACAATTTATCCCCCTCATTTAACTGGAAGAAATTTATGGATGATTCTAAAATTAGCAA
GTTCATGAACGAATTGGGAGAAATGGGCTATAAATTCCAATTCATTACTCTTGCAGGFTTGGCACTTA
ATTAACTACCATACATTCAAATTAGCGAGAGCATATAGAAATGAAGGAATGCCAGCATATGTAAGACT
ACAAGAGTTAGAGTTCCAAGCACAAGCAGAAGGCTATACTGCAGTAAGCCACCAGAGAGAGGTTGGAA
CAGATTACTTTGACCTAGTGTTAACAATAGCGTCTGGGGGAGAAGCGTCCACTACTGCAATGGAAGGT
TCCACTGAAGCTGAACAGTTTGTAGAGGCAAAACAGAAAGTATTAAAATAG

1c. Glyoxylate reductase – (annotated as a putative phosphoglycerate dehydrogenase)

agtgttctgt**tatattaaaag**attaaaacctatgaagtaaagtagggttt**ATGAAGATAATATCTACA**
GAAAAAGTTCCGGATGAGTGTAAAAATATAATCAGTGTTAAAGATGAGAATATAACGGAAGAGGATTA
TAAAAATGCAGAGATCTTGCTAACGTGGCCTGGTAGAGTAAATAACGATCTGATAGGTAAGATGCCTA
ATCTTAAGGTAATACAAACGTTTTCGGCGGGAGTTGATGATTTAGACTTCTCTATAATTCCTTCTCAT
GTTAAAGTATTCTCAAATGCGGGTGCTTACTCTTTGTCTGTAGCTGAACACGCTTGGGCTCTAATTTT
AGCATCAGCCAAAGGTGTAGGAACAAAAAAGAGGACTATAGTTTATGATGTTTCTGAAAAAACGTTGT
TGATTTTAGGTGGTGGGGGGATAGGATCTGAAGTAGCAAGGATAGGTAAAACAGCATTTAGGAATTAC
GTAATAGGTATTTCCAGATCTTTCAAGAAACCTGAATGGTTTGACGAAAGGCATGGAATGACTATGCT
AAGGGAGAAAATTGGAGAGGCTGATATAATTGTAGATACGCTACCTCTAAACAAACAAACTAGGGGTC
TTTTAAGTTATGATCTACTTAAAGATATAAGGCGTAATGCCATAATTGTTAATGTTGGGAGAGGTGAG
ACTGTGGATGAAGAAGGTATTTATAAGTTACTTAAGGAGAGGCAAGACGTAAGGTTTGGGACTGATGT
GTTTTGGAGAAAAAACGGTAAGGAGGATTTCTATAATACGAAGCTTTGGGAATTAGATAACTTTACTG
GAACTCTACACACCGCTGGGGCTTATGGCAATGAGAGTATAATGAAGAGGGCAATGCTCATAGCTTGT
TTAAATGTGAAAAAATATATAGATAGGGGAATTGCAGATAATGAAGTAAGAAGGGAAGATTATGTTTG
A

1d. Aldehyde Oxidoreductase (three subunits)

cattttattacctaatattcctagctaaacatttttacgtgataaaataaatagctaagagtgcaat
ataatatGTGTATCCGCCAGATTTTACATATGTTAGGGTTAGTAGCTCTGAGGAAGCTACAAAATTTC
TAGAATCCCACGACGATGCAAGGCCTCTAGCTGGAGGACAAAGTTTAATTCCAATGCTTAAACTTCGT
GTAATATCGCCCAATTATATAGTTGACCTAAATCCTATAACGTCACTAAGTTATGTAAGAAGTTCCTT
TAATTCAACTAAAATTGGTGCTCTAACTCGATATAATGAAATACTAAAGAATGATCTAGTAAGGGTAA
ACGTTCCATTACTTCATCAAGCAGTTAGGGTAGTAGGAGATATGCAAGTTAGAAACTTAGGTACTATT
GGTGGTAGCGCTGCAAACGCTGATCCATCAGCTGATATCCCCACTGTACTTACTGCGTTAAACGCCGA
AATTATTCTATCCTCAGCATCCGGCAATAGATCAGTTAATGCTCTAGATTTCTTTAAAGGCGCATTCG
CCACAGATTTAAGAAAAGGTGAAATTATCTCTGAAATTGTTTTACCTAACTTGGAGGGATATAGAACA
ATTTACAAAAAGGTCGTAAGAAGAGCTGGAGATTTTGCACTTGTATCTCTAGCATTAGCGATAAAATT
GAGGCAAAATGAGATAGAGGATATTAGATTAGCTTATGGTGGAGTTGGGGAGAGACCATTCAGAGCGT
TAGAAGTTGAGAAAAGTGTAATGGGTAAAAGGCTAAATGATGAGTTAGTAGAGGAAATTGTAAGTAAG
GTTTCAAGTCAAGTAAATCCCCCTTCCGATACTAGGGGGAGTTCTTGGTATAGGAGGGAGGTTATGAA
GGTTATAACTAGAAAAGCCTTAAAGGAGGTGTCGGGTTAAATGTTAATGGTTAATCAAGGAGAGAAAA
TTAAAATAAAAGTTAAGGTAAATGGAGTATTGTATGAGAGATATGTAAGTCCAAGGATACTATTAGTT
GACTTTTTAAGGGAAGAGTTAGGCTTAACGGGAACGAAAATTGGATGTGATACTACAACTTGTGGCGC
ATGCACAGTTCTATTAAATGGTAAATCAGTAAAATCTTGTACATTGTTTGCAGTGCAAGCTGATGGTG
CAGAAATAACTACAATTGAGGGTCTCTCAGTAGATTCTAAGCTTCATCCAATTCAAGAGGCGTTTAAG
GAAAATTTCGCCCTTCAGTGTGGTTTCTGTACGCCGGGAATGATAATGCAGGCATATTTCTTGCTGAA
AGAAAATCCAAATCCTTCTGAAGAGGAGGTTAGAGATGGACTTCACGGTAACATTTGTAGGTGTACTG
GATATCAAAATATTGTTAAGGCAGTTTTAGATGCTTCAAGGAGGTTGAGAGCATGAGCTACGTTGGCA
AGCCAGTAAAAAGAATTTATGACGATAAATTCGTAACAGGTAGAAGCACTTATGTTGACGATATAAGG
ATACCAGCTCTATATGCCGGTTTTGTTAGAAGTACTTATCCCCATGCAATTATTAAGAGAATTGATGT
TAGTGATGCGTTAAAGGTTAATGGTATAGTTGCAGTATTTACTGCAAAGGAAATTAACCCCTTATTAA
AAGGTGGAATTAGACCTTGGCCTACGTATATAGATATACGATCATTTAGGTACAGTGAGAGGAAGGCA
TTTCCAGAAAATAAGGTCAAATATGTAGGTGAACCAGTAGCAATTGTGCTTGGCCAAGACAAGTATAG
CGTTAGAGACGCCATAGATAAGGTAGTAGTTGAGTATGAACCCTTAAAACCAGTTATTAGAATGGAAG
AAGCAGAAAAAGATCAAGTAATAATCCATGAGGAGTTAAAAACTAACATATCCTATAAGATTCCATTT
AAAGCAGGAGAAGTTGACAAGGCATTCAGTGAATCTGATAAGGTAGTTAGAGTTGAGGCCATAAATGA
AAGATTAATTCCAAATCCCATGGAACCTAGAGGAATTGTATCTAGGTTTGAAGCAGGAACACTATCAA
TATGGTACTCTACTCAAGTTCCTCATTATATGCGTTCAGAATTTGCCAGGATACTTGGTATACCCGAA
AGTAAGATAAAAGTCAGTATGCCGGATGTGGGAGGTGCCTTTGGGGCTAAAGTTCATTTAATGCCAGA
AGAGTTAGCAGTAGTAGCATCGTCAATCATTTTAGGGAGGCCAGTAAGATGGACAGCTACTAGAAGTG
AGGAGATGTTAGCCAGTGAAGCGAGGCATAACGTTTTCACTGGTGAGGTAGCAGTTAAAAGAGATGGT
ACTATTCTAGGTATAAAGGGTAAACTATTGCTAGATCTAGGAGCCTATATAACAGTAACTGCCGGTAT
CCAACCATTAATAATACCAATGATGATACCCGGCCCCTACAAAATACGTAATCTGGATATCGAAAGCG
TTGCAGTCTACACTAATACCCCACCAATTACTATGTACAGAGGAGCCAGTAGACCAGAGGCAACATAT
ATAATTGAAAGGATAATGAGTACAGTGGCTGACGAGTTAGGGTTAGATGATGTAAGTATTAGGGAAAA
GAATCTAGTCACTGAATTACCATATACAAATCCGTTTGGTTTAAGGTATGATAGTGGAGACTATGTTG
GATTATTAAGAGAAGGCGTGAAGAGGTTAGGTTATTACGAACTTAAGAAGTGGGCTGAAGAGGAGAGA
AAGAAGGGGCATAGGGTTGGAGTAGGGTTAGCGTATTATCTGGAAATATGTAGTTTTGGTCCATGGGA
ATACGCTGAAGTTAGAGTGGATGAGAGGGGTGATGTATTAGTCGTTACTGGTACAACACCTCATGGAC
AAGGTACAGAAACTGCAATAGCTCAAATAGTTGCAGACGCCTTACAAATAGATATAAGTAGAGTTAGG
GTAATATGGGGAGATACTGATACTGTTGCAGCCAGTATGGGAACTTATGGTTCAAGATCTGTAACAAT
AGGTGGCTCTGCAGCAATTAAAGTTGCAGAAAAAATCTTAGATAAGATGAAGAGAATTGCAGCATCTA
CTTGGAATGTCGATGTTCAAGAAGTTCAATATGAGAAAGGAGAGTTTAAGTTAAAGAATGATCCAAGT
AAGAAGATGAGTTGGGACGACGTTGCTAGCATAGCGTATAGAAGTCATGATCCTGGGCTAGTAGAGAA
GATAATCTACGAGAACGATGTGACTTTCCCCTATGGAGTCCATATAGCGACAGTGGAGGTAGATGATA
CTGGAGTTGCTAGAGTTTTAGAATATAGGGCATACGATGATATTGGGAAGGTTGTAAATCCAGCATTA
GCAGAAGCGCAGATTCATGGTGGAGGTGTCCAGGCTGTTGGACAAGCACTATATGAGCAAGCTTTACT
CAATGAGAATGGACAGTTAATCGTAACTTATGCAGATTATTATGTCCCAACTGCTGTTGAGGCGCCTA
AGTTTACGTCAGTCTTTGCTGATCAATATCATCCATCTAACTATCCAACTGGTAGTAAGGGTGTTGGA
GAGGCAGCATTAATCGTAGGTCCTGCAGTGATAATTAGAGCATTGGAGGACGCTATTGGCACTAGATT
CACTAAAACTCCAACTACTCCAGAGGAAATTTTAAGGGCTATCGCTAGTAAAAGATGA

Appendix 2 – Sequencing of xylonate dehydratase (XD) gene in pMZ1 and pMJ0503 vectors

2a. Sequencing of XD gene in pMZ1 using vector specific primers (see Materials and Methods). Restriction sites are shown in colour (*Nco*1 – Red, *Bam*H1 – Green) and underlined. Strep tag and His tag with corresponding amino acid sequences are shown in grey and dark red respectively.

```
XD_PMZ1         TGTTNAACAGTTAGGTATACTATTTATAAAATAGTTGGGTCATAAAAGTACCCGAGACCATGG
XD_Seq          ------------------------------------------------------------A

XD_PMZ1         TGACAAAGATTTCAGAAATCGAAGCGTACATTTTAGGTAAGGAAGTAACAAGCGCTCAATGGG
XD_Seq          TGACAAAGATTTCAGAAATCGAAGCGTACATTTTAGGTAAGGAAGTAACAAGCGCTCAATGGG

XD_PMZ1         CTTCATTAATGGTACTGGTAAGGGTAACCACAAACGATGGTAGAGTAGGCTGGGGAGAAACTG
XD_Seq          CTTCATTAATGGTACTGGTAAGGGTAACCACAAACGATGGTAGAGTAGGCTGGGGAGAAACTG

XD_PMZ1         TAAGTGCCTTAAGGGCAGAAGCTGTGGCAAATTTTGTAAAAAAGATAAATACTGTATTAAAAG
XD_Seq          TAAGTGCCTTAAGGGCAGAAGCTGTGGCAAATTTTGTAAAAAAGATAAATACTGTATTAAAAG

XD_PMZ1         GAAATGACGTGTTTAACGTTGAAAAGAACAGGCTAGAATGGTATAAACATGATTTCAATATGA
XD_Seq          GAAATGACGTGTTTAACGTTGAAAAGAACAGGCTAGAATGGTATAAACATGATTTCAATATGA

XD_PMZ1         CTATCTCCTTAGAGTCTACAACAGCCTATAGTGCAGTAGATATAGCTTCTTGGGACATAATCG
XD_Seq          CTATCTCCTTAGAGTCTACAACAGCCTATAGTGCAGTAGATATAGCTTCTTGGGACATAATCG

XD_PMZ1         GAAAAGAGTTAGGTGCTCCATTATATAAGCTGCTTGGAGGAAAAACAAGAGATAAAGTACTTG
XD_Seq          GAAAAGAGTTAGGTGCTCCATTATATAAGCTGCTTGGAGGAAAAACAAGAGATAAAGTACTTG

XD_PMZ1         TCTACGCTAATGGTTGGTATCAAAACTGCGTGAAGCCAGAGGACTTTGCGGAAAAAGCTAAGG
XD_Seq          TCTACGCTAATGGTTGGTATCAAAACTGCGTGAAGCCAGAGGACTTTGCGGAAAAAGCTAAGG

XD_PMZ1         AGATCGTAAAAATGGGATATAAGGCTTTGAAATTTGACCCATTTGGTCCGTACTTTAACGATA
XD_Seq          AGATCGTAAAAATGGGATATAAGGCTTTGAAATTTGACCCATTTGGTCCGTACTTTAACGATA

XD_PMZ1         TTTCAAAGAAAGGACTAGATATAGCTGAGGAGAGAGTAAAGGCTGTTAGAGAGGCAGTTGGAG
XD_Seq          TTTCAAAGAAAGGACTAGATATAGCTGAGGAGAGAGTAAAGGCTGTTAGAGAGGCAGTTGGAG

XD_PMZ1         ACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTCGGCTATTATGATAG
XD_Seq          ACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTCGGCTATTATGATAG

XD_PMZ1         CGAAAAGACTAGAGAAGTATAATCCATTGTTCATGGAAGAACCAATTCATCCTGAGGATGTGG
XD_Seq          CGAAAAGACTAGAGAAGTATAATCCATTGTTCATGGAAGAACCAATTCATCCTGAGGATGTGG

XD_PMZ1         AAGGACTTAGAAAATATAGGAACAATACGAGCTTAAGGATTGCGTTAGGTGAAAGGATAATTA
XD_Seq          AAGGACTTAGAAAATATAGGAACAATACGAGCTTAAGGATTGCGTTAGGTGAAAGGATAATTA

XD_PMZ1         AATAAACAGCAAGCCTTATACTTTATGAAAGAGGGATTAGTGGATTTCTTACAAGCTGATCTA
XD_Seq          AATAAACAGCAAGCCTTATACTTTATGAAAGAGGGATTAGTGGATTTCTTACAAGCTGATCTA

XD_PMZ1         TACAGAATTGGTGGAGTTACTGAGACCAAGAAGGTAGTAGGGATTGCTGAGACCTTTGACGTA
XD_Seq          TACAGAATTGGTGGAGTTACTGAGACCAAGAAGGTAGTAGGGATTGCTGAGACCTTTGACGTA

XD_PMZ1         CAGATGGCTTTTCACAATGCTCAAGGTCCAATACTAAACGCCGTAACACTACAGTTTGACGCA
XD_Seq          CAGATGGCTTTTCACAATGCTCAAGGTCCAATACTAAACGCCGTAACACTACAGTTTGACGCA

XD_PMZ1         TTTATACCAAACTTCTTAATACAAGAGTCTTTCTATGATTGGTTCCCGAGTTGGAAGAGGGAG
XD_Seq          TTTATACCAAACTTCTTAATACAAGAGTCTTTCTATGATTGGTTCCCGAGTTGGAAGAGGGAG

XD_PMZ1         CTAATATATAATGGTACTCCAATAGATAACGGATACGCTATAATACCGGAAAGACCGGGTTTA
XD_Seq          CTAATATATAATGGTACTCCAATAGATAACGGATACGCTATAATACCGGAAAGACCGGGTTTA

XD_PMZ1         GGAGTTGAAGTAAATGAGAAAATGTTAGATAGTTTAAAGGTTAAAGGTGAGGAGTACTTTAAT
XD_Seq          GGAGTTGAAGTAAATGAGAAAATGTTAGATAGTTTAAAGGTTAAAGGTGAGGAGTACTTTAAT

XD_PMZ1         CCAGAAGAACCAGTGTGGGTAGTTAAAGGAACTTGGAGAGACTATGGATCCGGATGGAGTCAT
XD_Seq          CCAGAAGAACCAGTGTGGGTAGTTAAAGGAACTTGGAGAGACTATTAG       W  S  H

XD_PMZ1         CCACAATTTGAGAAGCATCACCATCATCACCATCATCACCATCATGAGGGCCCACTCCAAGC
                 P  Q  F  E  K  H  H  H  H  H  H  H  H  H  H  H
```

2b. Sequencing of XD gene in pMJ05-03 using vector specific primers (see Materials and Methods). Restriction sites are shown in colour (*Nco*1 – red, *Avr*II – blue, *Bam*H1 – green, *Eag*1 - purple) and underlined. Strep tag and His tag with corresponding amino acid sequences are shown in grey and dark red respectively.

```
XD_pMJ0305_F      ACCTCCCCTTTTTCNGCNGATAAGTCTGNTGGGTNTTTTNNCAGCTCCCCTAGGCACCNCNTGT
XD_SEQ            ------------------------------------------------------------

XD_pMJ0305_F      GTCGNGNTNTTTNTTANACGATCTTANANAACTCCGAGCTTCGATCTCTTATGTGCNTTGTGGT
XD_SEQ            ------------------------------------------------------------

XD_pMJ0305_F      CACTGTTGAATTTTCACGATCATCTANGGNCTCCCANAAACATACNTCATGTANCANNACATTA
XD_SEQ            ------------------------------------------------------------

XD_pMJ0305_F      ANTGAAANATAGANAANAGTTATATTATAGTTATTTTNAGAAAAACATCCAATATGTTAACAAA
XD_SEQ            ------------------------------------------------------------

XD_pMJ0305_F      ACGTCTTTTACGGAAATATATAAATGTTAAACANGTTAGGTATACTATTTATAAAATAGTTGGG
XD_SEQ            ------------------------------------------------------------

XD_pMJ0305_F      TCATAAAAGTACCCGAGACCATGGTGACAAAGATTTCAGAAATCGAAGCGTACATTTTAGGTAA
XD_SEQ            ----------------------ATGACAAAGATTTCAGAAATCGAAGCGTACATTTTAGGTAA

XD_pMJ0305_F      GGAAGTAACAAGCGCTCAATGGGCTTCATTAATGGTACTGGTAAGGGTAACCACAAACGATGGT
XD_SEQ            GGAAGTAACAAGCGCTCAATGGGCTTCATTAATGGTACTGGTAAGGGTAACCACAAACGATGGT

XD_pMJ0305_F      AGAGTAGGCTGGGGAGAAACTGTAAGTGCCTTAAGGGCAGAAGCTGTGGCAAATTTTGTAAAAA
XD_SEQ            AGAGTAGGCTGGGGAGAAACTGTAAGTGCCTTAAGGGCAGAAGCTGTGGCAAATTTTGTAAAAA

XD_pMJ0305_F      AGATAAATACTGTATTAAAAGGAAATGACGTGTTTAACGTTGAAAAGAACAGGCTAGAATGGTA
XD_SEQ            AGATAAATACTGTATTAAAAGGAAATGACGTGTTTAACGTTGAAAAGAACAGGCTAGAATGGTA

XD_pMJ0305_F      TAAACATGATTTCAATATGACTATCTCCTTAGAGTCTACAACAGCCTATAGTGCAGTAGATATA
XD_SEQ            TAAACATGATTTCAATATGACTATCTCCTTAGAGTCTACAACAGCCTATAGTGCAGTAGATATA

XD_pMJ0305_F      GCTTCTTGGGACATAATCGGAAAAGAGTTAGGTGCTCCATTATATAAGCTGCTTGGAGGAAAAA
XD_SEQ            GCTTCTTGGGACATAATCGGAAAAGAGTTAGGTGCTCCATTATATAAGCTGCTTGGAGGAAAAA

XD_pMJ0305_F      CAAGAGATAAAGTACTTGTCTACGCTAATGGTTGGTATCAAAACTGCGTGAAGCCAGAGGACTT
XD_SEQ            CAAGAGATAAAGTACTTGTCTACGCTAATGGTTGGTATCAAAACTGCGTGAAGCCAGAGGACTT

XD_pMJ0305_F      TGCGGAAAAAGCTAAGGAGATCGTAAAAATGGGATATAAGGCTTTGAAATTTGACCCATTTGGT
XD_SEQ            TGCGGAAAAAGCTAAGGAGATCGTAAAAATGGGATATAAGGCTTTGAAATTTGACCCATTTGGT

XD_pMJ0305_R      CCGTACTTTAACGATATTTCAAAGAAAGGACTAGATATAGCTGAGGAGAGAGTAAAGGCTGTTA
XD_SEQ            CCGTACTTTAACGATATTTCAAAGAAAGGACTAGATATAGCTGAGGAGAGAGTAAAGGCTGTTA

XD_pMJ0305_R      GAGAGGCAGTTGGAGACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTC
XD_SEQ            GAGAGGCAGTTGGAGACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTC

XD_pMJ0305_R      GGCTATTATGATAGCGAAAAGACTAGAGAAGTATAATCCATTGTTCATGGAAGAACCAATTCAT
XD_SEQ            GGCTATTATGATAGCGAAAAGACTAGAGAAGTATAATCCATTGTTCATGGAAGAACCAATTCAT

XD_pMJ0305_R      CCTGAGGATGTGGAAGGACTTAGAAAATATAGGAACAATACGAGCTTAAGGATTGCGTTAGGTG
XD_SEQ            CCTGAGGATGTGGAAGGACTTAGAAAATATAGGAACAATACGAGCTTAAGGATTGCGTTAGGTG

XD_pMJ0305_R      AAAGGATAATTAATAAACAGCAAGCCTTATACTTTATGAAAGAGGGATTAGTGGATTTCTTACA
XD_SEQ            AAAGGATAATTAATAAACAGCAAGCCTTATACTTTATGAAAGAGGGATTAGTGGATTTCTTACA

XD_pMJ0305_R      AGCTGATCTATACAGAATTGGTGGAGTTACTGAGACCAAGAAGGTAGTAGGGATTGCTGAGACC
XD_SEQ            AGCTGATCTATACAGAATTGGTGGAGTTACTGAGACCAAGAAGGTAGTAGGGATTGCTGAGACC

XD_pMJ0305_R      TTTGACGTACAGATGGCTTTTCACAATGCTCAAGGTCCAATACTAAACGCCGTAACACTACAGT
XD_SEQ            TTTGACGTACAGATGGCTTTTCACAATGCTCAAGGTCCAATACTAAACGCCGTAACACTACAGT

XD_pMJ0305_R      TTGACGCATTTATACCAAACTTCTTAATACAAGAGTCTTTCTATGATTGGTTCCCGAGTTGGAA
XD_SEQ            TTGACGCATTTATACCAAACTTCTTAATACAAGAGTCTTTCTATGATTGGTTCCCGAGTTGGAA

XD_pMJ0305_R      GAGGGAGCTAATATATAATGGTACTCCAATAGATAACGGATACGCTATAATACCGGAAAGACCG
XD_SEQ            GAGGGAGCTAATATATAATGGTACTCCAATAGATAACGGATACGCTATAATACCGGAAAGACCG

XD_pMJ0305_R      GGTTTAGGAGTTGAAGTAAATGAGAAAATGTTAGATAGTTTAAAGGTTAAAGGTGAGGAGTACT
XD_SEQ            GGTTTAGGAGTTGAAGTAAATGAGAAAATGTTAGATAGTTTAAAGGTTAAAGGTGAGGAGTACT
```

```
XD_pMJ0305_R      TTAATCCAGAAGAACCAGTGTGGGTAGTTAAAGGAACTTGGAGAGACTATGGATCCGGATGGAG
XD_SEQ            TTAATCCAGAAGAACCAGTGTGGGTAGTTAAAGGAACTTGGAGAGACTAT          W

XD_pMJ0305_R      TCATCCACAATTTGAGAAGCATCACCATCATCACCATCATCACCATCATTGAGGGCCCACTTTC
XD_SEQ            S H  P  Q  F  E  K  H  H  H  H  H  H  H  H  H  H

XD_pMJ0305_R      TCAAGTCTCACTATACCAANTGAGTTTTCTTTTANTCTTATTCTAGATGCTCATAAAAANTCGN

XD_pMJ0305_R      NTNGCCGGCCGCCACCNCGGGGAGCNAGNAAAAAACNAATGCATGTCATAGCTTTTCCNNNNAA
```

Appendix 3: Crystal screen conditions of recombinant malate synthase from *S. solfataricus*

Malate synthase was purified to homogeneity as described and concentrated to ~8mg/ml. Crystal screens JCSG+96 and Clear Strategy II (Molecular Dimensions, Suffolk UK) listed below were set up using the sitting drop method with three protein concentrations in each condition (1:1, 2:1 and 3:1 protein to screen) in a total of 300 – 450nl. Trays were incubated at 15°C and checked every 24h for crystal growth.

3a. JCSG screen

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
A1	4.5	0.2	lithium sulphate	0.1M sodium acetate	5-% v/v	PEG 400
A2	5.5			0.1M citrate	20% w/v	PEG 3K
A3		0.2	di-ammonium hydrogen citrate		20% w/v	PEG 3350
A4	4.6	0.02	calcium chloride	0.1M sodium acetate	30% v/v	MPD
A5		0.2	magnesium formate		20% w/v	PEG 3350
A6	4.2	0.2	lithium sulphate	0.1M phosphate / 0.1M citrate	20% w/v	PEG 1K
A7	9.5			0.1M CHES	20% w/v	PEG 8K
A8		0.2	ammonium formate		20% w/v	PEG 3350
A9		0.2	ammonium chloride		20% w/v	PEG 3350
A10		0.2	potassium formate		20% w/v	PEG 3350
A11	8.5	0.2	ammonium dihydrogen phosphate	0.1M tris	50% v/v	MPD
A12		0.2	potassium nitrate		20% w/v	PEG 3350
B1	4.0			0.1M sodium citrate	0.8M	ammonium sulphate
B2		0.2	sodium thiocyanate		20% w/v	PEG 3350
B3	9.0			0.1M bicine	20% w/v	PEG 6K
B4	7.5			0.1M HEPES	10% w/v	PEG 8K / 8% v/v Ethylene glycol
B5	6.5			0.1M sodium cacodylate	40% v/v	MPD / 5% w/v PEG 8K
B6	4.2			0.1M phosphate / 0.1M citrate	40% v/v	Ethanol 5% w/v PEG 1K
B7	4.6			0.1M sodium acetate	8% w/v	PEG 4K
B8	7.0	0.2	magnesium chloride	0.1M tris	10% w/v	PEG 8K
B9	5.0			0.1M sodium citrate	20% w/v	PEG 6K
B10	6.5	0.2	magnesium chloride	0.1M sodium cacodylate	50% v/v	PEG 200
B11	6.5				1.6M	tri-sodium citrate
B12		0.2	tri-potassium citrate		20% w/v	PEG 3350
C1	4.2	0.2	sodium chloride	0.1M phosphate / 0.1M citrate	20% w/v	PEG 8K
C2	4.0	1	lithium chloride	0.1M Na citrate	20% w/v	PEG 6K
C3		0.2	ammonium nitrate		20% w/v	PEG 3350
C4	7.0			0.1M HEPES	10% w/v	PEG 6K
C5	7.5			0.1M sodium HEPES	0.8M	sodium dihydrogen phosphate
C6	4.2			0.1M phosphate / 0.1M citrate	0.8M	potassium dihydrogen phosphate
C7	4.5	0.2	zinc acetate	0.1M sodium acetate	40% v/v	PEG 300
C8	8.5			0.1M tris	10% w/v	PEG 3K
C9	6.2			0.1M sodium phosphate / 0.1M potassium phosphate	20% v/v	Ethanol
C10	9.0			0.1M bicine	25% v/v	1,2-propanediol
C11	4.6			0.1M sodium acetate	10% v/v	Glycerol
C12					10% w/v	PEG 20K / 2% v/v Dioxane
D1					2M	ammonium sulphate

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
D2	7.5	0.2	magnesium chloride	0.1M sodium HEPES	10% w/v	PEG 1K / 10% w/v PEG 8K
D3	6.2	0.2	sodium chloride	0.1M sodium phosphate / 0.1M potassium phosphate	24% w/v	PEG 1.5K / 20% v/v Glycerol
D4	4.5	0.2	lithium sulphate	0.1M sodium acetate	30% v/v	PEG 400
D5	7.5			0.1M HEPES	50% v/v	PEG 200
D6	8.5	0.2	magnesium chloride	0.1M tris	30% w/v	PEG 8K
D7	8.5	0.2	lithium sulphate	0.1M tris	70% v/v	MPD
D8	8.0			0.1M tris	20% w/v	PEG 8K
D9		0.17	ammonium sulphate		40% v/v	PEG 400
D10	6.5	0.2	calcium acetate	0.1M sodium cacodylate	40% v/v	MPD
D11	4.6	0.14	calcium chloride	0.1M sodium acetate	25.5% w/v	PEG 4K / 15% v/v Glycerol
D12		0.04	potassium dihydrogen phosphate		40% v/v	PEG 300
E1	6.5			0.1M sodium cacodylate	14% v/v	2-propanol / 30% v/v Glycerol
E2	6.5	0.2	sodium chloride	0.1M sodium cacodylate	16% w/v	PEG 8K / 20% v/v Glycerol
E3	7.5	0.2	sodium chloride	0.1M HEPES	1M	tri-sodium citrate
E4	8.5	0.2	lithium sulphate	0.1M tris	2M	ammonium sulphate
E5	10.5			0.1M CAPS	10% v/v	2-propanol
E6	8.0	0.2	zinc acetate	0.1M imidazole	1.26M	ammonium sulphate
E7	6.5	0.2	zinc acetate	0.1M sodium cacodylate	40% v/v	MPD
E8	4.5			0.1M sodium acetate	20% w/v	PEG 3K
E9	6.5			0.1M MES	10% v/v	2-propanol
E10	9.0			0.1M bicine	1M	di-ammonium hydrogen phosphate
E11	6.5	0.16	calcium acetate	0.1M sodium cacodylate	1.6M	magnesium sulphate
E12	8.0			0.1M imidazole	10% w/v	PEG 6K
F1	6.5		caesium chloride	0.1M MES	14.4% w/v	PEG 8K / 20% v/v Glycerol
F2	5.0			0.1M sodium citrate	10% w/v	PEG 8K
F3	8.0			0.1M tris	30% v/v	Jeffamine M-600
F4	7.5			0.1M HEPES	3.2M	ammonium sulphate
F5	8.5	0.2	magnesium chloride	0.1M tris	20% v/v	MPD
F6	9.0			0.1M bicine	20% v/v	Jeffamine M-600
F7	7.0				50% v/v	ethylene glycol
F8	7.0				10% v/v	MPD
F9	7.0				0.8M	succinic acid
F10	7.0	1.1	sodium malonate	0.1M HEPES	2.1M	DL-malic acid
F11	7.0	1	succinic acid	0.1M HEPES	2.4M	sodium malonate
F12	7.0			0.1M HEPES	0.5% v/v	Jeffamine ED-2001
G1	7.0			0.1M HEPES	1% w/v	PEG 2K MME
G2	7.5	0.02	magnesium chloride	0.1M HEPES	30% v/v	Jeffamine M-600
G3	8.5	0.01	cobalt chloride	0.1M tris	30% v/v	Jeffamine ED-2001
G4	8.5	0.2	tri-methylamine N-oxide	0.1M tris	22% w/v	polyacrylic acid 5100 sodium salt
G5	7.5	0.005	cobalt chloride / cadmium chloride / magnesium chloride / nickel chloride	0.1M HEPES	20% w/v	polyvinylpyrrolidone K15
G6	7.0	0.2	sodium malonate		20% w/v	PEG 2K MME
G7	7.0	0.1	succinic acid		12% w/v	PEG 3350
G8	7.0	0.15	DL- malic acid		20% w/v	PEG 3350
G9		0.1	potassium thiocyanate		15% w/v	PEG 3350
G10		0.15	potassium bromide		20% w/v	PEG 3350
G11	5.5			0.1M Bis Tris	30% w/v	PEG 2K MME
G12	5.5			0.1M Bis Tris	30% w/v	PEG 2K MME
H1	5.5			0.1M Bis Tris	2M	ammonium sulphate
H2	5.5	1	ammonium sulphate	0.1M Bis Tris	3M	sodium chloride
H3	5.5			0.1M Bis Tris	0.3M	magnesium formate

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
H4	5.5	0.2	calcium chloride	0.1M Bis Tris	1% w/v	PEG 3350
H5	5.5	0.2	ammonium acetate	0.1M Bis Tris	25% w/v	PEG 3350
H6	5.5	0.1	ammonium acetate	0.1M Bis Tris	45% v/v	MPD
H7	5.5	0.2	ammonium sulphate	0.1M Bis Tris	45% v/v	MPD
H8	5.5	0.2	sodium chloride	0.1M Bis Tris	17% w/v	PEG 10K
H9	5.5	0.2	lithium sulphate	0.1M Bis Tris	25% w/v	PEG 3350
H10	5.5	0.2	ammonium acetate	0.1M Bis Tris	25% w/v	PEG 3350
H11	5.5	0.2	magnesium chloride	0.1M Bis Tris	25% w/v	PEG 3350
H12	7.5	0.2	ammonium acetate	0.1M HEPES	25% w/v	PEG 3350
					25% w/v	PEG 3350
					45% v/v	MPD

3b. Clear Strategy Screen

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
A1	5.5	1.5	ammonium sulphate	0.1M sodium acetate		
A2	5.5	0.8	lithium sulphate	0.1M sodium acetate		
A3	5.5	2	sodium formate	0.1M sodium acetate		
A4	5.5	0.5	potassium dihydrogen phosphate	0.1M sodium acetate		
A5	5.5	0.2	calcium acetate	0.1M sodium acetate	25% w/v	PEG 2K MME
A6	5.5	0.2	calcium acetate	0.1M sodium acetate	15% w/v	PEG 4K
A7	5.5	2.7	ammonium sulphate	0.1M sodium acetate		
A8	5.5	1.8	lithium sulphate	0.1M sodium acetate		
A9	5.5	4	sodium formate	0.1M sodium acetate		
A10	5.5	1	potassium dihydrogen phosphate	0.1M sodium acetate		
A11	5.5	0.2	calcium acetate	0.1M sodium acetate	10% w/v	PEG 8K / 10%w/v PEG 1K
A12	5.5	0.2	calcium acetate	0.1M sodium acetate	8% w/v	PEG 20K / 8% v/v PEG 550 MME
B1	5.5			0.1M sodium acetate	40% v/v	MPD
B2	5.5			0.1M sodium acetate	40% v/v	1,4-Butanediol
B3	5.5	0.01	cadmium chloride	0.1M sodium acetate	20% w/v	PEG 4K
B4	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	20% v/v	PEG 550 MME
B5	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	20% v/v	PEG 600
B6	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	20% w/v	PEG 1.5K
B7	5.5			0.1M sodium acetate	35% v/v	2-Propanol
B8	5.5			0.1M sodium acetate	30% v/v	Jeffamine M-600
B9	5.5	0.01	nickel chloride	0.1M sodium acetate	20% w/v	PEG 4K
B10	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	18% w/v	PEG 3350
B11	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	18% w/v	PEG 5K MME
B12	5.5	0.15	potassium thiocyanate	0.1M sodium acetate	15% w/v	PEG 6K
C1	6.5	1.5	ammonium sulphate	0.1M sodium cacodylate		
C2	6.5	0.8	lithium sulphate	0.1M sodium cacodylate		
C3	6.5	2	sodium formate	0.1M sodium cacodylate		
C4	6.5	0.5	potassium dihydrogen phosphate	0.1M sodium cacodylate		
C5	6.5	0.2	calcium acetate	0.1M sodium cacodylate	25% w/v	PEG 2K MME
C6	6.5	0.2	calcium acetate	0.1M sodium cacodylate	15% w/v	PEG 4K
C7	6.5	2.7	ammonium sulphate	0.1M sodium cacodylate		
C8	6.5	1.8	lithium sulphate	0.1M sodium cacodylate		
C9	6.5	4	sodium formate	0.1M sodium cacodylate		
C10	6.5	1	potassium dihydrogen phosphate	0.1M sodium cacodylate		

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
C11	6.5	0.2	calcium acetate	0.1M sodium cacodylate	10% w/v	PEG 8K / 10%w/v PEG 1K
C12	6.5	0.2	calcium acetate	0.1M sodium cacodylate	8% w/v	PEG 20K / 8% v/v PEG 550 MME
D1	6.5			0.1M sodium cacodylate	40% v/v	MPD
D2	6.5			0.1M sodium cacodylate	40% v/v	1,4-Butanediol
D3	6.5	0.01	cadmium chloride	0.1M sodium cacodylate	20% w/v	PEG 4K
D4	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	20% v/v	PEG 550 MME
D5	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	20% v/v	PEG 600
D6	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	20% w/v	PEG 1.5K
D7	6.5			0.1M sodium cacodylate	35% v/v	2-Propanol
D8	6.5			0.1M sodium cacodylate	30% v/v	Jeffamine M-600
D9	6.5	0.01	nickel chloride	0.1M sodium cacodylate	20% w/v	PEG 4K
D10	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	18% w/v	PEG 3350
D11	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	18% w/v	PEG 5K MME
D12	6.5	0.15	potassium thiocyanate	0.1M sodium cacodylate	15% w/v	PEG 6K
E1	7.5	1.5	ammonium sulphate	0.1M Tris		
E2	7.5	0.8	lithium sulphate	0.1M Tris		
E3	7.5	2	sodium formate	0.1M Tris		
E4	7.5	0.5	potassium dihydrogen phosphate	0.1M Tris		
E5	7.5	0.2	calcium acetate	0.1M Tris	25% w/v	PEG 2K MME
E6	7.5	0.2	calcium acetate	0.1M Tris	15% w/v	PEG 4K
E7	7.5	2.7	ammonium sulphate	0.1M Tris		
E8	7.5	1.8	lithium sulphate	0.1M Tris		
E9	7.5	4	sodium formate	0.1M Tris		
E10	7.5	1	potassium dihydrogen phosphate	0.1M Tris		
E11	7.5	0.2	calcium acetate	0.1M Tris	10% w/v	PEG 8K / 10%w/v PEG 1K
E12	7.5	0.2	calcium acetate	0.1M Tris	8% w/v	PEG 20K / 8% v/v PEG 550 MME
F1	7.5			0.1M Tris	40% v/v	MPD
F2	7.5			0.1M Tris	40% v/v	1,4-Butanediol
F3	7.5	0.01	cadmium chloride	0.1M Tris	20% w/v	PEG 4K
F4	7.5	0.15	potassium thiocyanate	0.1M Tris	20% v/v	PEG 550 MME
F5	7.5	0.15	potassium thiocyanate	0.1M Tris	20% w/v	PEG 600
F6	7.5	0.15	potassium thiocyanate	0.1M Tris	20% w/v	PEG 1.5K
F7	7.5			0.1M Tris	35% v/v	2-Propanol
F8	7.5			0.1M Tris	30% v/v	Jeffamine M-600
F9	7.5	0.01	nickel chloride	0.1M Tris	20% w/v	PEG 4K
F10	7.5	0.15	potassium thiocyanate	0.1M Tris	18% w/v	PEG 3350
F11	7.5	0.15	potassium thiocyanate	0.1M Tris	18% w/v	PEG 5K MME
F12	7.5	0.15	potassium thiocyanate	0.1M Tris	15% w/v	PEG 6K
G1	8.5	1.5	ammonium sulphate	0.1M Tris		
G2	8.5	0.8	lithium sulphate	0.1M Tris		
G3	8.5	2	sodium formate	0.1M Tris		
G4	8.5	0.5	potassium dihydrogen phosphate	0.1M Tris		
G5	8.5	0.2	calcium acetate	0.1M Tris	25% w/v	PEG 2K MME
G6	8.5	0.2	calcium acetate	0.1M Tris	15% w/v	PEG 4K
G7	8.5	2.7	ammonium sulphate	0.1M Tris		
G8	8.5	1.8	lithium sulphate	0.1M Tris		
G9	8.5	4	sodium formate	0.1M Tris		
G10	8.5	1	potassium dihydrogen phosphate	0.1M Tris		
G11	8.5	0.2	calcium acetate	0.1M Tris	10% w/v	PEG 8K / 10%w/v PEG 1K
G12	8.5	0.2	calcium acetate	0.1M Tris	8% w/v	PEG 20K / 8% v/v PEG 550 MME
H1	8.5			0.1M Tris	40% v/v	MPD

	pH	Conc (M)	Salt	Buffer	Units	Precipitant
H2	8.5			0.1M Tris	40% v/v	1,4-Butanediol
H3	8.5	0.01	cadmium chloride	0.1M Tris	20% w/v	PEG 4K
H4	8.5	0.15	potassium thiocyanate	0.1M Tris	20% v/v	PEG 550 MME
H5	8.5	0.15	potassium thiocyanate	0.1M Tris	20% v/v	PEG 600
H6	8.5	0.15	potassium thiocyanate	0.1M Tris	20% w/v	PEG 1.5K
H7	8.5			0.1M Tris	35% v/v	2-Propanol
H8	8.5			0.1M Tris	30% v/v	Jeffamine M-600
H9	8.5	0.01	nickel chloride	0.1M Tris	20% w/v	PEG 4K
H10	8.5	0.15	potassium thiocyanate	0.1M Tris	18% w/v	PEG 3350
H11	8.5	0.15	potassium thiocyanate	0.1M Tris	18% w/v	PEG 5K MME
H12	8.5	0.15	potassium thiocyanate	0.1M Tris	15% w/v	PEG 6K

Appendix 4: Alignment of archaeal malate synthase sequences ordered by optimum growth temperature

The archaeal malate synthase sequences are from *Pyrobaculum arsenaticum* (100°C), *Pyrobaculum aerophilum* (95°C), *Pyrobaculum calidifontis* (95°C), *Pyrobaculum islandicum* (90°C), *Thermoproteus tenax* (90°C), *Sulfolobus acidocaldarius* (75°C), *Sulfolobus solfataricus* (75°C). Residues were selected when there appeared to be a conservative pattern of change with increasing growth temperature (highlighted in yellow). Key catalytic residues identified in *E. coli* are either fully conserved (highlighted in black) or semi conserved (highlighted in grey).

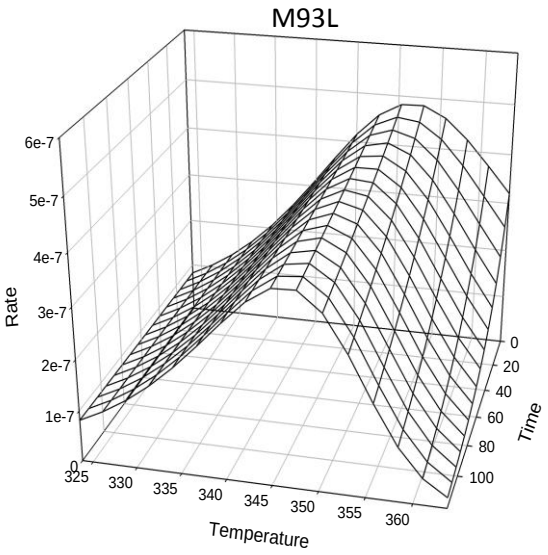
```
Pyrobaculum arsenaticum      ---MIEISPSVLREYGDLFGEKTVNGRRISVEGLIEELTRELRAEIDRVI
Pyrobaculum aerophilum       ---MIEINPLVVKNYGDLLGVKTVNNRQVSVEGLIEELVREFRDEINRVI
Pyrobaculum calidifontis     ---MIEINPKVLHLYRDLFGEKVVNGRRVVVEQLIEDLTREFRAEIDRVI
Pyrobaculum islandicum       --MVIEINREVLKKFPELFGEKIVNGRKVVVEDLIEKLTREFRSDIDRVV
Thermoproteus tenax          --MVIEINKEVLKKFPELFGEKTVNGRRVVVEELIERLTRELRAEIDKAV
Sulfolobus acidocaldarius    MPSKLKIPDEIYHNYKDLFGEKVINDRIVSVEGLIEELSVEFSDEIKRVI
Sulfolobus solfataricus      MSSSLKIPEEIYQKYSDLFGEKTINDRIVSVEKLIEELAVEFSDEIRKVI
                                 ::*   : : : :*:* * :*.* : ** *** *  *:  :* :.:
                                                                      M93
Pyrobaculum arsenaticum      RARREWLNDKRPLKLKAAFPSWEEKFTDADGNVRTFREIVQGLIDNLLGR
Pyrobaculum aerophilum       RARREWLNDKRPVRVKGAFPHWEEKFVDADGNVRTFREIVQGLIDNLLDR
Pyrobaculum calidifontis     RARREWLERREPVRVKAAFPKWDDKFVDADGNVRTFREIVQGLIDNFLGR
Pyrobaculum islandicum       RARREWLNDRRPVREKATFPRWDEKFVDADGNVRTFREIVQGLIDNFLGR
Thermoproteus tenax          RARREWLNDRRPVREKAAFPQWDQRFVDADGNVRTFREIVQGLIDNFLGR
Sulfolobus acidocaldarius    SKRRKWLESKDSVSSKGSFPSWDEVFVDADGNRRTFREIVQGMIDNFLQR
Sulfolobus solfataricus      NKRRLWLESKDPVTSKGAFPSFDEVFVDADGNKRTFREIIQGMIDNFLGV
                               ** **: : .:  *.:** ::: *.***** ******:**:***:*

                           Q101/S102       I114                M135 N140
Pyrobaculum arsenaticum      DTPLRWGLNWNTPVPDDLHPLKNPGLEITGPWSPMSRAIHQINADVASMM
Pyrobaculum aerophilum       DSPLRWGLNWNTPVPDDLHPLKNPGLEITGPWYPMSRAIHQINADVASMM
Pyrobaculum calidifontis     DTPLRWGLNWNTPVPDDLHPLKNPGLEITGPWFPMSRAIHQINADVASMM
Pyrobaculum islandicum       DTPLRWGLNWNTPVPDDLHPLKNPGLEITGPWYPMSRAIHQINADVASMM
Thermoproteus tenax          DTPLRWGLNWNTPVPDDLHPLKNPGLEITGPWHPMSRAIHQINADVVSMM
Sulfolobus acidocaldarius    QSSLSWRLNDHVPIPNDAHPIKNPGLEITGPWYPLSRAYNQVNTDVTSAM
Sulfolobus solfataricus      QSKLKWRLNENVPIPKDAHPLNNPGLEITGPWYPLSRAYNQINSDVACVM
                             ::  * * ** :.*:*.* **::********** *:*** :*:*:**.. *

Pyrobaculum arsenaticum      EDEEDASPAWYIPRGSGRTTAAVWEARRIVNRVLRGDVPQPYYEGGKEYR
Pyrobaculum aerophilum       EDEEDASPAWYIPRGSGRAVAAVWEARRVVNRVLRGEVPQPYYEGGKEYR
Pyrobaculum calidifontis     EDEEDASPAWYVPWGSGREVAPVWEARRIVKRVLSGDVPDPYYEGGKVYR
Pyrobaculum islandicum       EDEEDASPAWFIPWGSGRSVAAVWEARRIVKRVLEGDIPTPYIEGGKAYY
Thermoproteus tenax          EDEEDASPAWFIPRGSGRTVAAVWEARRIVKRVLEGDIPTPYVEGGKTYY
Sulfolobus acidocaldarius    EDEEDASPAWYTPYGSGKSIADVWDARKNVKLFLSGKAPNPYYEKGKTYT
Sulfolobus solfataricus      EDEEDASPAWYIPFGSGKTIADVWEGRKNVKLFLSGKAPNPYYEKGKTYS
                             **********: * ***:  * **:.*: *: .* *. * ** * ** *

Pyrobaculum arsenaticum      IKKPREKWPTLIHRVPGLHILDFDIRVDGNPVPAIITSVVIYTVNNYDLL
Pyrobaculum aerophilum       IKKPRDKWPTLIHRVPGLHILDFDIRVDGKPAPAIITSIVIYTVNNYDLL
Pyrobaculum calidifontis     IQKPRERWPTLIHRVPGLHILDFDVRVDGRPVPAIITSIVIYTVNNYDAL
Pyrobaculum islandicum       VKKERSKWPTLIHRIPGLHILDFDIRVDGRPVPAVITSIVIYTVNNYDQL
Thermoproteus tenax          LKKERSKWPTLIHRVPGLHILDFDIRVDGRPAPAIITSIVIYTVNNYEAL
Sulfolobus acidocaldarius    LNKPRDKWPTIFHRLPGLHLLDFDITLDGKPVPAIIVSAVIYTLNNYISL
Sulfolobus solfataricus      LNKPRDKWPVIFHRLPGLHLLDFDITLNGKPVPAIIVSAVIYTLNNYNSL
                             ::* *.:**.::**:****:****: ::*.*.**:*.* ****:***  *
```

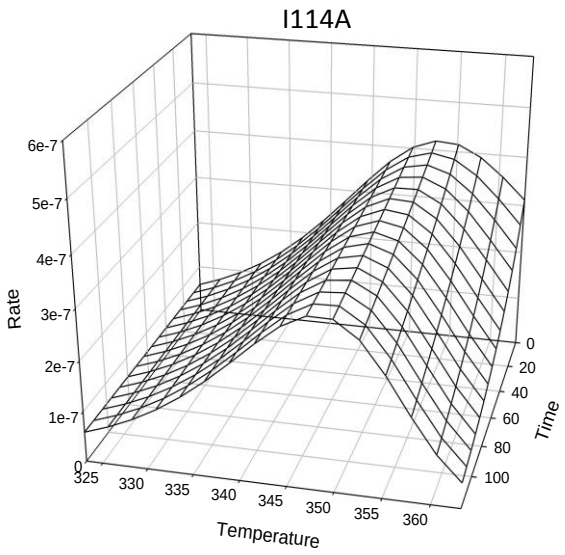
```
                                                                L291
Pyrobaculum arsenaticum       KRAGSGVYFYVPKVQTPDEALVVEKLLRRVEDKLGLRRGELKIAMLYEEA
Pyrobaculum aerophilum        KRAGSGVYFYIPKVQTPDEALVIEKILRRVEDRLGLKRGELKIAMLYEEA
Pyrobaculum calidifontis      KRAGSGVYFYVPKVQTPEEALVVEKILRRVEDALGLRRGELKIAMLYEEA
Pyrobaculum islandicum        KKVGSGVYFYVPKVQTPDEALVIEKILRRVEDELGLKRGETKIAMLYEEA
Thermoproteus tenax           KRAGSGVYFYVPKTQTPEEALVIEKILRRVEDELGLRRGETKLAMLYEEA
Sulfolobus acidocaldarius     RNAGSGVYFYLPKTQTPEEALIVEKILRRIEEKLGLPIGTLKLALLYEEV
Sulfolobus solfataricus       KSAGSGVYFYLPKTQTPDEALIIEKILRRIESKLGLKIGTLKLALLYEEV
                              : .*******:**.***:***::**:***:*. ***  * :*:*:****.

Pyrobaculum arsenaticum       RAGLYLPVIFWIWRERLVKSNNGRWDYLGSLIEMWKDEAVYPDPQNITMT
Pyrobaculum aerophilum        RAGLYLPVIFWIWRERLVKSNNGRWDYLGSLIEMWKDEAVYPDPQNITMT
Pyrobaculum calidifontis      RAGLYLPVIFWIWRERLVKSNNGRWDYLGSLIEMWKDEAVYPDPQNITMT
Pyrobaculum islandicum        RAGLYLPVIFWIWRERLVKSNNGRWDYLGSLIEMWKDEAVYPDPQNITMT
Thermoproteus tenax           RAGLYLPVIFWIWRERLVKSNNGRWDYLGSLIEMWKDEAVYPDPQNITMT
Sulfolobus acidocaldarius     NAGRYFPVILWIFRERLIKSNNGRWDYLGSLIEMWLQERVLPDPQNITMT
Sulfolobus solfataricus       NAGRFFPVILWIFRERLIKSNNGRWDYLGSLIEMWLQEKVLPDPQNITMT
                              .** ::***:**:****:**************** :* * *********

Pyrobaculum arsenaticum       HPVMMAYQKWNALMCLMAGLDRQGKLNAGPVGGMAAVMLYRPDDPYQRHR
Pyrobaculum aerophilum        HPIMMAYQKWNALLCLMAGLDRGGRLNAAPVGGMAAVMLYRPDDPYQRNR
Pyrobaculum calidifontis      HPIMMAYQKYNALMCLMAGLGRDGRLNAAPVGGMAAVMLYRPDDPYQRHR
Pyrobaculum islandicum        HPIMMAYQKYNALLCLMAGLDREGKLNAAPVGGMAAVMLYRPDDPYQRNR
Thermoproteus tenax           HPIMMAYQKYNALLCLMAGLDREGKLNAGPVGGMAAVMLYRQDDPYQRNR
Sulfolobus acidocaldarius     SPNMMAYQKYNALIMLLAGAK-NGEADAAPVGGMAAVMLYPQSDAFERHR
Sulfolobus solfataricus       SPNMMAYQKYNALMMLLAGAK-NGEADAAPVGGMAAVMLYPQTDPFGRNR
                               * ******:***: *:**    *. :*.***********   *.: *:*

Pyrobaculum arsenaticum       FNQRALRAIWLDKLRERLIGLIFVTEEPVK-KVTLRDVLEGKVKGRLFDL
Pyrobaculum aerophilum        FNARALRAIWLDKLRERLIGLIFLTEEPVR-KVTLNDILKGRVKGRLFDL
Pyrobaculum calidifontis      YNARALRAIWLDKLRERLIGLIFVTEEPVK-KVKLREILEGKVRGRLYDL
Pyrobaculum islandicum        YNARALRAIWLDKLRERLIGLIFVTEEAVS-KVTLKDILEGKVKGRLYDL
Thermoproteus tenax           YNARALRAIWLDKLRERLIGLVFVTEDAVG-KVTLDDVLKGRVRGRLYDL
Sulfolobus acidocaldarius     YNLKALRGIKLDKLRERLIGLIFIGD--VKGKVTLEDILKGKVQGKLYDM
Sulfolobus solfataricus       YNLKALRGIKLDKLRERLIGLIFVAEGKVEGKITLEDIVNGKVKGKLYDM
                              :* :***.* ***********:*: :  *  *:.* ::::*:*:*:*:*:

                                                 P469
Pyrobaculum arsenaticum       FRQSWVATPEESYVKAGNEPLRASLEELQQMINRPVKFVEVDGVKIPTVD
Pyrobaculum aerophilum        FRQSWVATPEESYVKAGNEPLRASLEELQQLINRPVKFVEVDGVKIPAVD
Pyrobaculum calidifontis      FRQSWVATPEEAYVRAGNEPLRARLEELEKLINRPVKYVEVDGVKIPAVD
Pyrobaculum islandicum        FRQSWVATPEESYVKAGNEPLKASLEELQAMINRPVKYVEVDSVKIPAVD
Thermoproteus tenax           FRQSWVATPEESYVKAGNEPLKASLEELQAMINRPVKFVEVDGVKIPAVD
Sulfolobus acidocaldarius     FRQSWVATKEEAYVEAGNKPLKASLEELQKMIDAPVEYVTVEGTKMPTVD
Sulfolobus solfataricus       FRQSWVATKEEAYVEAGSKPLRASLDELQKMIDAPINYIEVEGTKLPTVD
                              ******** **:**.**.:*:* *:**: :*: *:::: *:..*:*:**

Pyrobaculum arsenaticum       SGLTEQERQLFIRLGLLDEQGNITPWVIRPDMLDTPEKLLGNPELWGGKD
Pyrobaculum aerophilum        SGLTEQERQLFIKLGLIDEDGNITPWVIRPDMLDTPEKLLNNPELWGGKD
Pyrobaculum calidifontis      SGLTEQERQLFIRLGLIDEAGNITPWVIRPDMVDTPEKLLINPELWGGAD
Pyrobaculum islandicum        SGLTEQERQLFQRLGLIDENGNITPWVIRPEMLDSPEKLFNNVELWGGKD
Thermoproteus tenax           SGLTEQERQLFQRLGLIDENGNITPWVIRPDMLDSPEKLFNNRELWGGRD
Sulfolobus acidocaldarius     SGLTPEERSLFQRLGLIDENGKITPWVITKDMIDTPEKLLHNKELWGGKS
Sulfolobus solfataricus       SGLTPEERALFQKLGLIDERGKITPWVITKEMINTPEKLLFNKELWGGKD
                              **** :** ** :***:** *:****** :*::*:****: * ***** .

Pyrobaculum arsenaticum       LWTALFEPPKGDITAEHIQHAFYMAANYGFQLLNGNLAAAIDDYELGQRF
Pyrobaculum aerophilum        LWSALFEPPKGDITAEHIQHAFYMAANYGFQLLNGNLAAAIDDYELGQRF
Pyrobaculum calidifontis      LWTALYQPPKGDITAEHIQHAFYMAANYGFQLLNGNLAAAIDDYELGQRF
Pyrobaculum islandicum        LWSALYEPPKGDITIEHIQHAFYMAANYGFQLLNGNLAAAIDDYELGQRF
Thermoproteus tenax           LWSALYEPPKGDITIEHIQHAFYMAANYGFQLLNGNLAAAIDDYELGQRF
Sulfolobus acidocaldarius     IWNALYDIPDGEITPEHVQHAFYMAANYGFQLLNGNLAAAIDDYELKQRF
Sulfolobus solfataricus       LWHSLYDIPEGDITPEHVQHAFYMAANYGFQLLNGNLAAAIDDYELKQRF
                              :* :*:: *.*:** **:**************************** ***
```

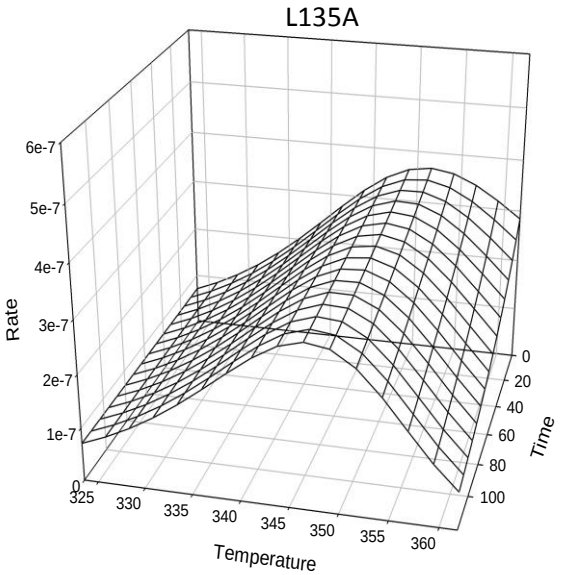
```
Pyrobaculum arsenaticum       MNDLATYRIFSTWLWTLLRHNAVITKDGAFKGPARTGLGVIPAEDRVKVA
Pyrobaculum aerophilum        MNDLATYRIFSTWLWTLLRHKAVITKDGAFKGPARTSLGVIPAEDRVKIP
Pyrobaculum calidifontis      MNDLATYRIFSTWLWTLLRHKAPITKDGAFKGPAKTALGVIPAEDRVKVA
Pyrobaculum islandicum        MNDLATYRIFSTWLWTLLRHKARITKDGALKGPAKTRLGVIPADDRIKIS
Thermoproteus tenax           MNDLATYRIFSTWLWTLLRHKARITKDGALKGPARTHLGVIPAEDRVRVP
Sulfolobus acidocaldarius     MNDLATYRIFTSWLWSLVNRDAVITKEGYIKAPKLTRDGVIPAENVLKVA
Sulfolobus solfataricus       MNDLATYRIFTSWLWSVINRDASFTKDGYIKGPKLTKDGVIPAEDVLKVT
                              **********::***:::.:.* :**:* :*.*  *  *****:: ::::.

Pyrobaculum arsenaticum       AGTRFTEELFDKLWDLHMEWTLAFYEDLDRIAAERILHRF--------VN
Pyrobaculum aerophilum        AGTPFTEELFDKLWDLHMEWTLAFYDDLDIIAAERILTRF--------TD
Pyrobaculum calidifontis      AGTPFTEELFDKLWELHMEWTWAFYDDLDRIASERILHKF--------VN
Pyrobaculum islandicum        AGTQFDEELFEKLWELHMEWTYAFYEDLDRISAERILLRF--------VE
Thermoproteus tenax           AGAQFTEELFEKLWDLHMEWTYAFYDDLDRISAERILLRF--------TE
Sulfolobus acidocaldarius     KGTK-VKDIFEEIWKLHFDWTNEFYKEQDMRVANRLAETFGKTNNNSVVE
Sulfolobus solfataricus       KGTK-VKDIFEKLWELHLDWTYEFYKEQDMRASRRIAETFGKTNNASTVE
                               *:   :::*:::*.**::**  **.: *   :.*:   *        .:

Pyrobaculum arsenaticum       RVRSAVAEAYKAGPFRYQSPRDTAKKIAESITVE--ELERAVVENQPRFD
Pyrobaculum aerophilum        RVKRAIAEAYRAGPFRYQSPRETAKKIAEVITEE--ELEKAVVENQPRFD
Pyrobaculum calidifontis      AVKEAVRDAYAAGPFRRVKPGDAARRVLSAIRAE--EVEKAVVENAPRFD
Pyrobaculum islandicum        NVKNILHNAYKAGPFRFQTPIDTARKIAELFKVE--ELEKAVIENQPRFD
Thermoproteus tenax           SVKGALREAYKAGPFRFQSPAETAKRISESIKAE--DLERAVVENQPRFD
Sulfolobus acidocaldarius     EIYNIISKAYNAGPFRQMSAKEAAEKIGKLLNTQPSKIEEELIKLAPRFD
Sulfolobus solfataricus       EVYKVISKAYSSGPFREMSAKEAAQKLAKILNANASEIEEELINLAPRFD
                               :   : .** :****  .. ::*.:: . :  :  .:*. :::  ****

                                                 I769   L776
Pyrobaculum arsenaticum       RSFAPVIMEILRAKLKSPMYLQHGGRLIMALAPLPDEERDAVLRAIFSPR
Pyrobaculum aerophilum        RSFAPVIMEILKTKLKSPMYLQHSGRLIMALAPLPDEERDLALKALFTPR
Pyrobaculum calidifontis      RAFAPVIMEILRAKLKSPMYLQHGGRLIMALAPLPDEVRDAALKAVFMPR
Pyrobaculum islandicum        RSFASVIMDILKVKLTSPMYLQHGGRLIMVLAPLPDEERSTVLRALFTPR
Thermoproteus tenax           RSFAPVIMDILRAKLTSPMYLQHSGRLIMALAPLPDEERGAVLQAMFSPR
Sulfolobus acidocaldarius     REMAPVIMEILMRQFLYPKYIMNSGKILFVLSPLDPEKRLKVMDSIFSFR
Sulfolobus solfataricus       RSMAPVIMEILMKEMLYPKYIMNSGKILFILSPLDPERRSKVMDSIFSFR
                              * :*.***:**  ::  * *: :.*::::: *:**  * *  .: ::*  *

Pyrobaculum arsenaticum       EEVERLVKEGKLKPYALELYDYVHDVR-
Pyrobaculum aerophilum        EEVERLVREGRLRPQVLELYDYIHDNR-
Pyrobaculum calidifontis      EEVERLVASGRLGKEVLEIYDYVHDIR-
Pyrobaculum islandicum        EEVEKLVKEGKLKSYVLELYDYIHDIR-
Thermoproteus tenax           EEVERLVKEGRLKPYVLELYDYVYDIRP
Sulfolobus acidocaldarius     EMVQDKVKRGELDPTVLELYDYIYDDYK
Sulfolobus solfataricus       KMIEDKVKRGELDKWVLELYDYVYDNYW
                              : ::  *  *.*   .**:***::*
```

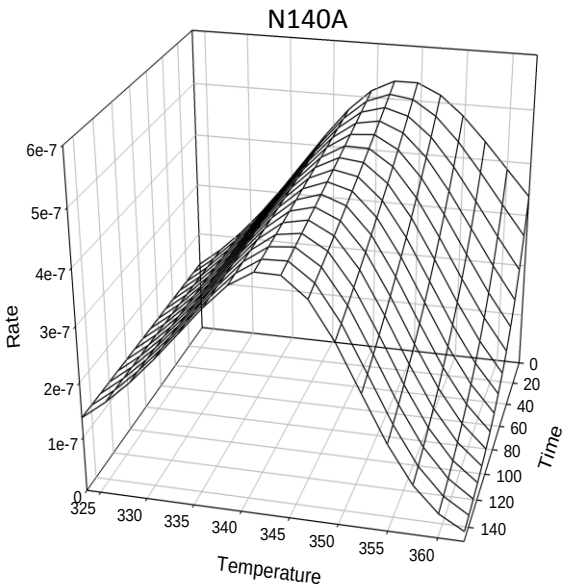
Appendix 5: Simulated 3D plots using parameters derived from the Equilibrium Model

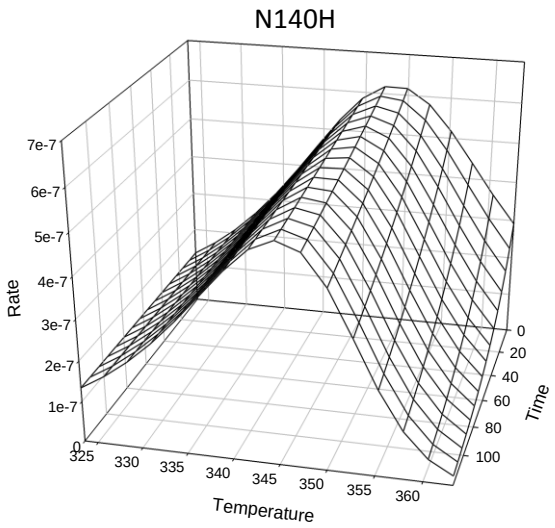
Simulated plots of rate vs time vs temperature using parameters derived from the equilibrium model. The graphs clearly show that all enzymes display an optimum temperature at time zero in accordance with the equilibrium model.

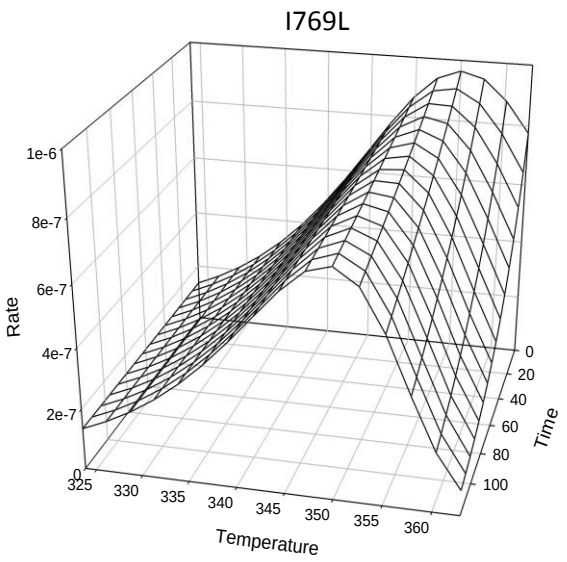


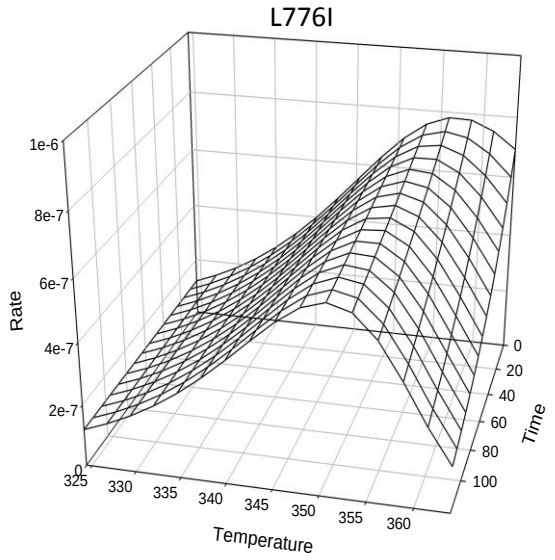


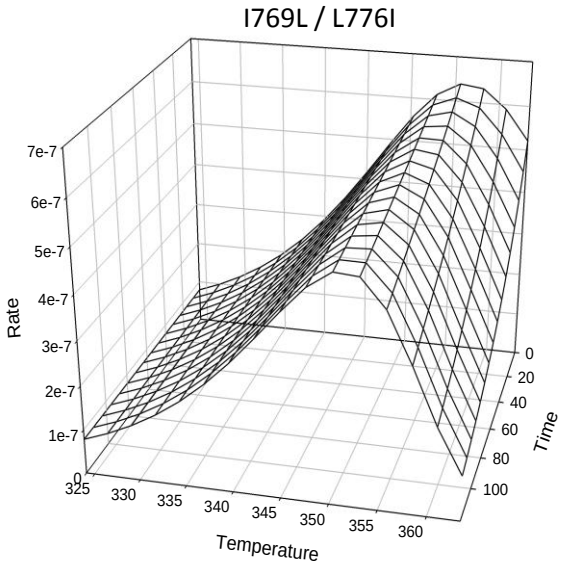


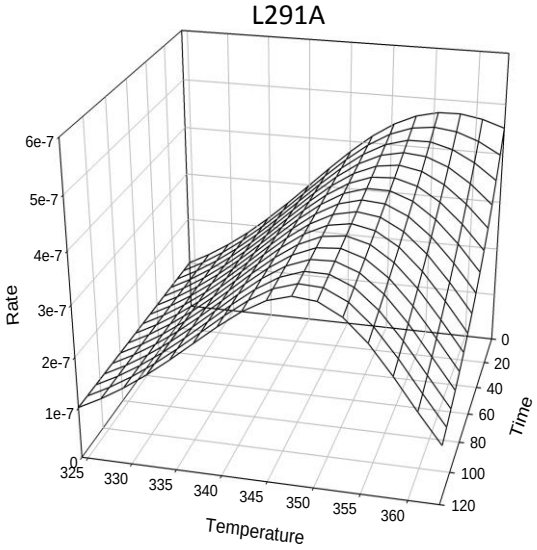


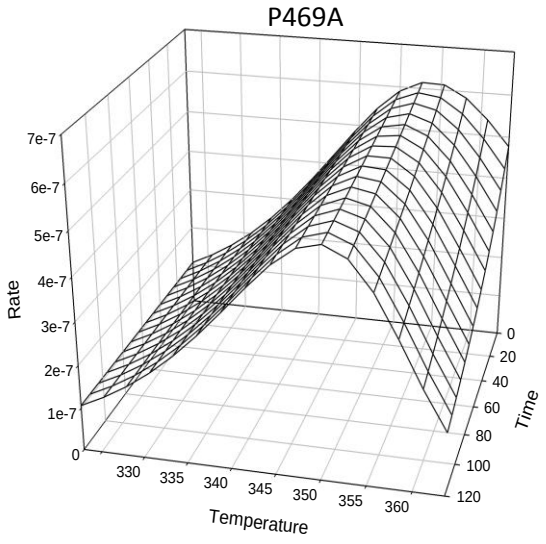

N140H
Rate
7e-7
6e-7
5e-7
4e-7
3e-7
2e-7
1e-7
0
325
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
Time


I769L
Rate
1e-6
8e-7
6e-7
4e-7
2e-7
0
325
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
Time


L776I
Rate
1e-6
8e-7
6e-7
4e-7
2e-7
0
325
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
Time


I769L / L776I
Rate
7e-7
6e-7
5e-7
4e-7
3e-7
2e-7
1e-7
0
325
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
Time

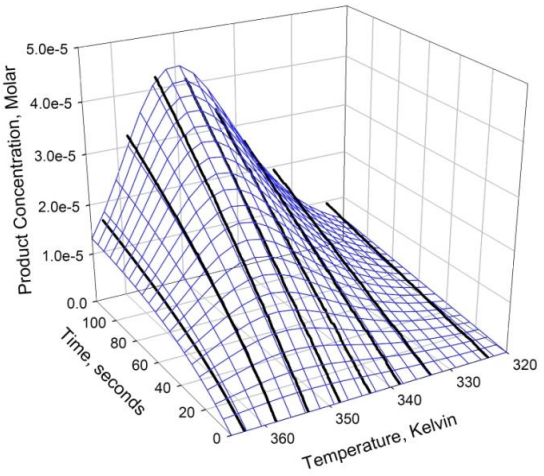

L291A
Rate
6e-7
5e-7
4e-7
3e-7
2e-7
1e-7
0
325
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
120
Time


P469A
Rate
7e-7
6e-7
5e-7
4e-7
3e-7
2e-7
1e-7
0
330
335
340
345
350
355
360
Temperature
0
20
40
60
80
100
120
Time

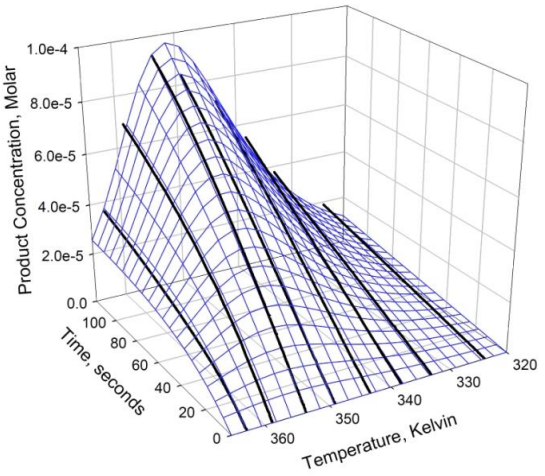
Appendix 6: Validation of final Equilibrium Model parameters

Parameters from the Equilibrium Model were used to generate a simulated 3D plot of increase in product concentration versus temperature versus time (blue mesh), which was then compared to the corresponding 2D plot generated from smoothed raw assay data (black lines). Parameters are shown as they are inputted into the MATLAB application; EAi = ΔG^*_{inact}, EAcat = ΔG^*_{cat}, HEq = ΔH_{eq} and TEq = T_{eq}.

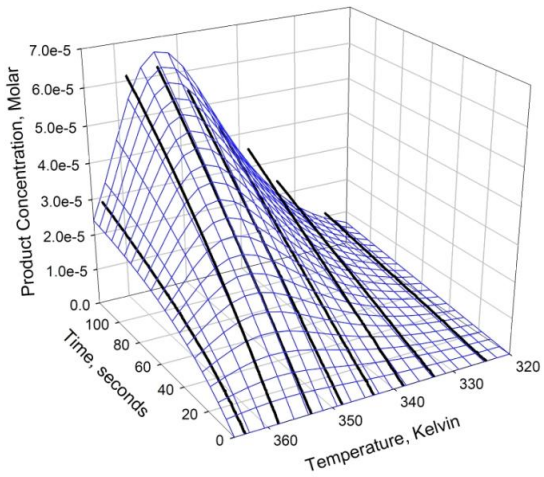
Malate Synthase I114A
Experimental and Matlab Simulation
E0 = 4.8228e-8
EAi = 101615.1226
EAcat = 78863.3002
HEq = 174300.3445
TEq = 354.0423



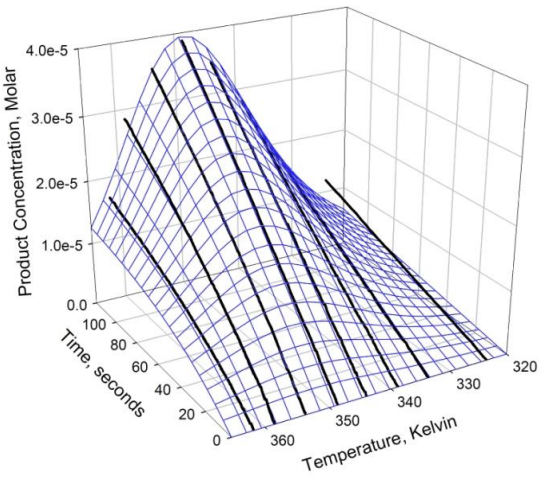
Malate Synthase I769L
Experimental and Matlab Simulation
E0 = 9.4681e-9
EAi = 100515.8126
EAcat = 72408.3533
HEq = 189357.0820
TEq = 357.5329



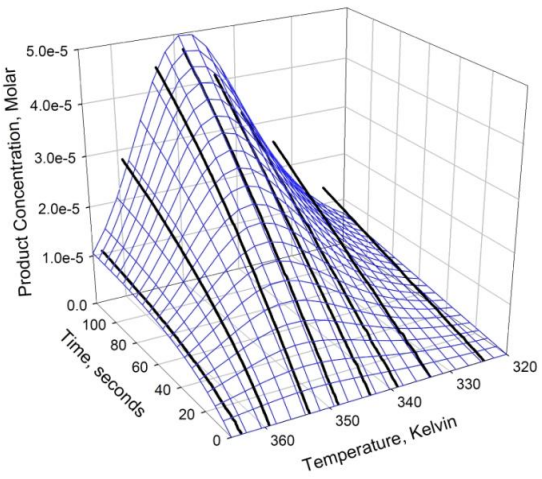
Malate Synthase I769L - L776I
Experimental and Matlab Simulation
E0 = 1.0532e-8
EAi = 101566.0864
EAcat = 74215.4053
HEq = 199170.2134
TEq = 358.7869



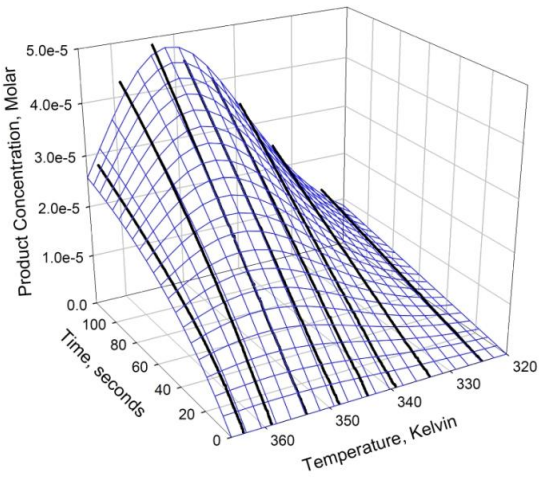
Malate Synthase L135A
Experimental and Matlab Simulation
E0 = 2.4557e-8
EAi = 102103.3245
EAcat = 76612.5136
HEq = 155764.3725
TEq = 351.2076



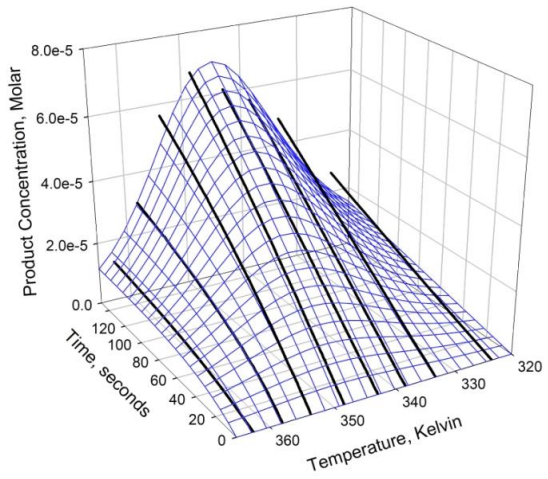
Malate Synthase M93L
Experimental and Matlab Simulation
E0 = 6.6929e-8
EAi = 100700.1027
EAcat = 79025.2020
HEq = 173866.2478
TEq = 352.4640



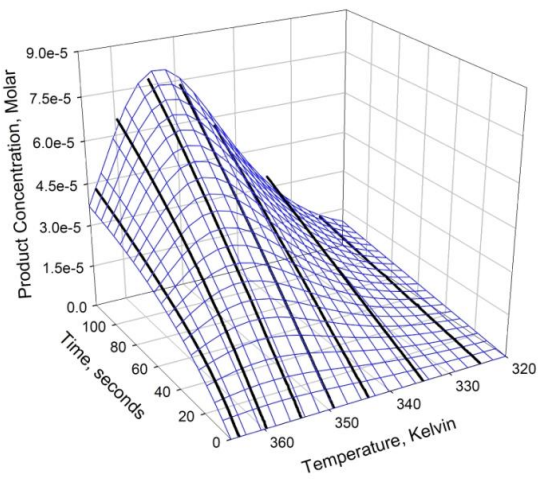
Malate Synthase L291A
Experimental and Matlab Simulation
E0 = 5.1489e-9
EAi = 102406.1919
EAcat = 71576.7657
HEq = 116666.1505
TEq = 351.9434



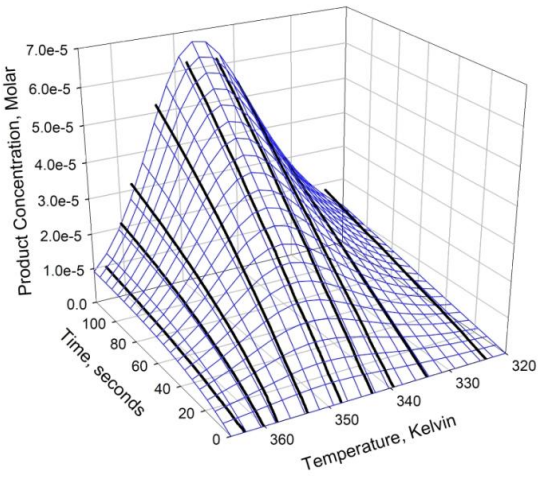
Malate Synthase N140A
Experimental and Matlab Simulation
E0 = 2.5910e-8
EAi = 101010.5123
EAcat = 74977.7541
HEq = 153630.5602
TEq = 348.1041



Malate Synthase L776I
Experimental and Matlab Simulation
E0 = 4.4326e-9
EAi = 101884.0881
EAcat = 70822.4027
HEq = 160031.5032
TEq = 358.0731



Malate Synthase N140H
Experimental and Matlab Simulation
E0 = 1.8898e-8
EAi = 100932.5570
EAcat = 74219.6733
HEq = 189384.7380
TEq = 350.5292



Malate Synthase P469A
Experimental and Matlab Simulation
E0 = 2.4468e-9
EAi = 102899.7123
EAcat = 69213.4356
HEq = 162090.1543
TEq = 355.2602